

University of Groningen

System design and wide-field imaging aspects of synthesis arrays with phased array stations

Bregman, Jacob Dirk

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

2012

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Bregman, J. D. (2012). *System design and wide-field imaging aspects of synthesis arrays with phased array stations: to the next generation of SKA system designers*. [Thesis fully internal (DIV), University of Groningen]. [s.n.].

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

RIJKSUNIVERSITEIT GRONINGEN

**System Design and Wide-field Imaging
Aspects of Synthesis Arrays with
Phased Array Stations**

To the next generation of SKA system designers

Proefschrift

ter verkrijging van het doctoraat in de
Wiskunde en Natuurwetenschappen
aan de Rijksuniversiteit Groningen
op gezag van de
Rector Magnificus, dr. E. Sterken,
in het openbaar te verdedigen op
vrijdag 14 december 2012
om 12.45 uur

door

Jacob Dirk Bregman

geboren op 13 juli 1946

te Amsterdam

Promotores: Prof. dr. W.N. Brouw
Prof. dr. H.R. Butcher

Beoordelingscommissie: Prof. dr. R.F. Peletier
Prof. dr. ir. A.J. van der Veen
Prof. ir. A. van Ardenne

ISBN: 978-90-367-5841-3
ISBN: 978-90-367-5840-6

(electronic version)

An aerial photograph of a large, circular artificial island situated in a river. The island is divided into several sections by narrow paths or ditches. Several clusters of solar panels are visible, arranged in a hexagonal pattern around the perimeter of the island. In the center of the island, there is a large field of small, square solar panels. The surrounding landscape is green and hilly, with a river flowing around the island.

System Design and Wide-field Imaging Aspects of Synthesis Arrays with Phased Array Stations

To the next generation of SKA system designers

Jaap D. Bregman

This dissertation covers the art of designing an aperture synthesis radio telescope like LOFAR. Well-known design principles are combined with a vision of new solutions that are expected to materialize in the near future, due to current technological developments. The central question is how scientific users with a given budget can achieve optimum results when the final instrument becomes operational

System design starts with an analysis of the fundamental limitations of image forming by means of aperture synthesis, and of practical limitations like the disturbance caused by the ionosphere. Such an analysis leads to the formulation of a number of scaling laws for the optimum array configuration, and for the amount of digital processing that will be required.

An aperture synthesis telescope consists of an array of stations, which can be of various types. In a phased-array type station, the signals from element antennas are added electronically in such a way that the sensitivity is maximized in a given direction. Phased-array technology offers the flexibility of distributing a given number of element antennas over an optimum number of stations.

Our research has shown that such stations must have a minimum size, to allow the effective correction of ionospheric disturbances over the entire field of view. Too small stations only allow proper correction for a limited part of their large field. Too few stations cause additional noise that can only partly be removed by more processing.

An important practical problem is the non-trivial amount of processing that is required for image forming. Therefore, an important result of this thesis is the development of new and efficient methods. Their processing is reduced to the theoretical minimum, i.e. proportional to the area of the field of view, expressed in resolution elements. For a sufficient number of stations in an optimal configuration, it should be possible to achieve minimum noise as well as minimum processing for the new generation of giant radio telescopes, from LOFAR to SKA and beyond.

*front picture: central core of LOFAR (Exloo, the Netherlands)
with 6 low-band stations and 12 high band phased arrays*

System Design and Wide-field Imaging Aspects of Synthesis Arrays with Phased Array Stations

To the next generation of SKA system designers

Dissertation by Jaap D. Bregman

Table of Contents	5
Preface	13
Glossary	19
1. Introduction	21
2. Overview of System Design for LOFAR	29
3. Efficient Processing for Wide-field Synthesis Imaging	55
4. Ionosphere Pathlength Variation and Self-Calibratability	175
5. Sensitivity limitations by Artefacts in Aperture Synthesis	257
6. Conclusions and Recommendations	295
Bibliography	311
Summary	319
Samenvatting	333
Dankwoord	349
Colofon	353

Table of Contents

Preface	13
Glossary	19
Introduction	21
2. Overview of System Design for LOFAR	29
2.1 Global design drivers for LOFAR	30
2.1.1 Design for Imaging	
2.1.2 Sensitivity types and impact on instrument design	
2.1.3 Minimum station size and calibratability	
2.1.4 Global Design Approaches	
2.1.5 Processing cost evolution over time	
2.2 LOFAR Characteristics	36
2.2.1 Array	
2.2.2 Stations	
2.2.3 Low Frequency issues and interference	
2.2.4 Signal processing at station and array level	
2.2.5 Field-of-View	
2.3 Calibration & imaging limitations at low frequencies	39
2.3.1 Sensitivity limits calibratability	
2.3.2 Image and source distortion relate to station and array size	
2.3.3 Array planarity, Field-of-View, and faceted imaging	
2.3.4 Intrinsic array planarity versus extrinsic baseline planarity	
2.3.5 Polarization correction in the image	
2.3.6 Deconvolution issues for synthesis imaging with a changing station beam	

2.4	Processing issues for imaging, correlation and beam-forming	45
2.4.1	Data output rate of correlation processing is a bottleneck for European LOFAR	
2.4.2	Correlation processing power as reference for image processing	
2.4.3	Processing for source subtraction and U,V-gridding Dominates over correlation processing	
2.4.4	Full Field-of-View can be handled in principle with dedicated imaging procedures	
2.4.5	Correlation on a general-purpose platform	
2.4.6	Dedicated station processing platforms versus general purpose correlation platform	
2.5	New Considerations in the Design of LOFAR	49
2.5.1	Short dipole	
2.5.2	Station configuration with expanding shells	
2.5.3	Calibratability, image forming & processing	
2.5.4	Grating lobes & blind angles	
2.5.5	FoV pattern of a snapshot image defined by the average over all station beams	
2.5.6	Snapshot corrections for beam shape and polarization	
2.5.7	Expo-shell array configuration	
2.5.8	Summary of paradigm shifts	
3.	Efficient Processing for Wide-field Synthesis imaging	55
3.1	Field-of-View of 2-D Fourier imaging with a nonplanar array	60
3.1.1	Basic Interferometer Measurement Equation	
3.1.2	3-D Fourier Inversion	
3.1.3	Spherical projection	
3.1.4	2-D Fourier inversion of Planar Array responses	
3.1.5	2-D Fourier inversion of data taken with a tilted array plane	

3.1.6	Phase after a fringe shift correction on correlated signals of a non-planar array	
3.1.7	Fringe stopping and fringe tracking	
3.1.8	Field-of-view limitation by non-planarity in 2-D Fourier imaging	
3.1.9	FoV for Intrinsic and extrinsic non-planarity	
3.1.10	Synthesis imaging with a single 2-D Fourier inversion	
3.1.11	Point Spread Function	
3.1.12	Combining direct and model based inversion to handle non-planarity	
3.1.13	Summary, Conclusions and main Result	
3.2	Decorrelation by averaging in frequency and time domain	80
3.2.1	Tolerated amplitude degradation	
3.2.2	Time averaging	
3.2.3	Frequency averaging	
3.2.4	Effects of the sinc shaped degradation function	
3.2.5	Correlation and post correlation processing impact	
3.3	Fast Fourier Transform imaging and filtering by Convolution	85
3.3.1	Resampling convolution of observed interferometer data	
3.3.2	Distortion correction by convolution	
3.3.3	Consequences for effective U,V-coverage of line and continuum observations	
3.4	Field-of-View extension of 2-D Fourier imaging with non-planar arrays	90
3.4.1	Quasi-convolution correction and W-projection	
3.4.2	Support of the quasi-convolution kernel	
3.4.3	Comparison with W-projection analysis and discussion	
3.4.4	Convolution processing determined by choice of U,V-reference plane	
3.4.5	Field-of-view of a 2-D Fourier image after complex quasi-convolution	
3.4.6	Fast Facet imaging	
3.4.7	Summary, Conclusions, and Results	

3.5	Snapshot synthesis in an array based coordinate system	107
3.5.1	2-D Snapshot imaging with a nonplanar array	
3.5.2	Sky tracking with a shifting correction for the 2-D Fourier image	
3.5.3	Duration of a synthesized snapshot observation	
3.5.4	Field rotation during sky tracking	
3.5.5	Synthesis imaging with synthesized snapshots	
3.5.6	How do sources outside the nominal FoV appear in a synthesis?	
3.5.6.1	Attenuation by side lobes of a phased array station	
3.5.6.2	Rotation and fringe track effects in synthesized snapshot imaging	
3.5.6.3	Summary and conclusions	
3.5.7	Summary and Results	
3.6	Phased array station beam aspects in synthesis imaging	123
3.6.1	Phase centre position of a phased array station	
3.6.2	Array element beam patterns and polarization characteristics	
3.6.3	Polarization of a phased array station beam	
3.6.4	Polarization over the phased array station beam after gain calibration	
3.6.5	Element beam pattern and blind angle effects	
3.6.6	Combining stations with different polarization characteristics	
3.6.7	Combining beams of stations with different diameter	
3.6.8	Summary and Conclusions	
3.7	Comparing processing for 3-D, 2-D and Synthesized snapshot imaging	145
3.7.1	Processing capacity of the main steps in hybrid imaging	149
3.7.1.1	Resolution and FoV determine number of visibility samples	
3.7.1.2	2-D FFT facet imaging	
3.7.1.3	Number of Facets and Size of the Convolution Kernel	
3.7.1.4	Fast Faceting	
3.7.1.5	Minimum number of convolution operations	
3.7.1.6	Number of source subtract operations	
3.7.1.7	Station beam and polarization correction	
3.7.1.8	Interpolation on a sky image grid	
3.7.1.9	Number of synthesized 2-D FFT snapshots and number of planes in 3-D	

3.7.2	Balancing Convolution and source subtraction against FFT processing	159
3.7.2.1	Source Subtraction dominates over convolution and Fourier inversion	
3.7.2.2	Continuum versus line observing	
3.7.2.3	3-D and 2-D synthesized snapshot imaging alternatives	
3.7.3	Comparing post correlation processing with correlation processing	161
3.7.3.1	Continuum imaging	
3.7.3.2	Comparison with legacy imaging packages and their successors	
3.7.4	Results and Conclusions	
3.8	Summary, Results, and Recommendations	166
4.	Ionosphere Pathlength Variation and Self-Calibratability	175
4.1	Refraction Basics	178
4.1.1	Refractive index of a plasma	
4.1.2	Faraday rotation	
4.1.3	Refraction by a horizontal surface observed by tilted telescope and horizontal array	
4.1.4	Refraction by a wedge	
4.1.5	Refraction by a curved slab derived from pathlength differences	
4.1.6	Refraction by two curved slabs	
4.2	Refraction by Troposphere and Ionosphere	188
4.2.1	Large scale model of troposphere and ionosphere	
4.2.2	Refraction by large and medium scale wedges in the ionosphere	
4.2.3	Spherical refraction contributions by the troposphere	
4.2.3	Spherical refraction contribution by the ionosphere	
4.2.4	Summary and conclusions	
4.3	Ionosphere phase delay screen contributions	194
4.3.1	TID waves in lower ionosphere	

4.3.2	Kolmogorov turbulence model	197
4.3.2.1	1-D analysis of GPS track data to define variation over phase screen	
4.3.2.2	2-D tip-tilt correction and residual deviation over small area of phase screen	
4.3.2.3	Differential angular position shift within a station beam	
4.3.2.4	Relevant time scales	
4.3.3	Comparison with interferometer data	200
4.3.3.1	Differential phase gradients over a large aperture	
4.3.3.2	Differential angular position shift and associated source degradation	
4.3.3.3	Large scale TID and small-scale Kolmogorov Turbulence results	
4.3.4	Summary and conclusions	
4.4	Multi-source self-calibration approach	207
4.5	Angular density of sources as function of their flux and size	212
4.5.1	Introducing cumulative and differential source counts	
4.5.2	Analysis of source counts at 38 MHz, 151 MHz, 325 MHz and 1.4 GHz	
4.5.3	Source sizes at 20 cm and 90 cm and suitability as LOFAR calibrators	
4.5.4	Source properties below 1 mJy	
4.5.5	Deriving 1.4 GHz cumulative source count and frequency scaling formulae	
4.5.6	Conclusions	
4.6	Number of expected calibration sources per station beam	222
4.6.1	Sensitivity of LOFAR interferometers	
4.6.2	Number of sources per beam for self-calibration of ionosphere and beam shape	
4.6.3	Improving the spatial sampling for the delay screen	
4.6.4	Summary and conclusions	
4.7	From interferometer phase to station based TEC screen values	230
4.7.1	From interferometer phase to delay, TEC and phase unwrapping requirements	

4.7.2	Decomposing Interferometer delay and TEC into station based delay and TEC	
4.7.3	Large scale refraction effects	
4.7.4	Differential delay screen corrections using a peeling approach	
4.7.5	Accuracy of station based phase delays	
4.7.6	TEC screen construction by renormalization of station based direction dependent TEC	
4.7.7	Conclusions	
4.8	Simplified polynomial interpolation model for the delay screen	242
4.8.1	Lagrange interpolation	
4.8.2	Accuracy of 2 nd order Lagrange interpolation for a TID sine wave model	
4.8.3	Delay screen accuracy limitations by Kolmogorov Turbulence	
4.8.4	Matching station beam width and effective integration times	
4.9	Summary of limitations in TEC screen modelling by self-calibration	248
4.10	Main Conclusions	254
5.	Sensitivity limitations by Artefacts in Aperture Synthesis	257
5.1	Confusion aspects in a synthesis image	259
5.2	Side lobe level in wide band snapshot synthesis imaging	262
5.2.1	Array configuration	
5.2.2	Quasi-convolution effects by bandwidth and time integration	
5.2.3	Frequency averaging	
5.2.4	Time averaging	
5.2.5	Combined frequency and time averaging	
5.2.6	Effect of sources outside the main beam	
5.2.7	Combining snapshots in a synthesis image	
5.2.8	Minimum number of source subtractions	
5.2.9	Processing implications	

5.3	Side lobe noise after self-calibration and source subtraction	280
5.3.1	Errors in nominal side lobes by array element based complex gain errors	
5.3.2	Noise contributions by error side lobes	
5.3.3	Noise contribution by self-calibration	
5.3.4	Noise contribution by phase screen calibration	
5.3.5	Noise contribution by Kolmogorov evolution in the phase screen	
5.3.6	Noise contribution by image phase errors	
5.3.7	Averaging of independent snapshot images	
5.4	Summary and Conclusions	288
6.	Conclusions & Recommendations	295
6.1	Scaling laws in Fourier imaging	297
6.2	Limitations by self-calibration	301
6.3	System design of synthesis arrays	304
6.4	Recommendations for LOFAR	305
6.5	Recommendations for SKA-Low	306
6.6	Main Results	308
	Bibliography	311
	Summary	319
	Samenvatting	333
	Dankwoord	349
	Colofon	353

Preface

Addressing the Big Questions of astronomy will require (among other things) a new generation of giant radio telescopes in the coming decades. Their antennas and receiving systems must be more sensitive, by some two orders of magnitude, than those of existing instruments, and offer an angular resolution of better than one second of arc. The former requires a very large collecting area and therefore the development of cheap, highly optimized signal chains. The latter requires that the collecting area be distributed over an array of widely separated “stations”, with baselines up to hundreds of km.

This was the subject of a workshop organized in Delft by ASTRON and CSIRO, directly following the URSI General Assembly in The Hague in August, 1996. At this workshop, it was already abundantly clear that a collecting area of about a million square meters (i.e. a square km) would be required [Ardenne, 1997]. Serious activity on the system design of this new generation of synthesis arrays started after a subsequent workshop, held in Sydney in December 1997. One of the themes was the use of phased arrays as an attractive alternative for the more traditional parabolic reflector antennas.

Phased arrays have considerable advantages; the most important one is providing the largest collecting area at giving cost, at least at low frequencies. This is why radio astronomy actually started with such arrays. Since they do not require a complex mechanical structure, or moving parts to track moving objects, they are much faster while cost is proportional to collecting area. In addition, they can survey the sky simultaneously in many different directions at a limited electronic cost increase.

However, phased arrays have some issues too. They have a spectral operating range that is limited not only by the properties of the constituent antenna elements, but also by array-effects such as grating lobes and blind angles of which the latter are caused by mutual coupling between the elements. In addition, the response beams of sparse arrays have high side-lobes, making them more sensitive for contaminating sources outside the field of interest. For all these reasons, dishes and phased arrays are complementary solutions, with dishes being more attractive for higher observing frequencies. The crossover point is likely to move upwards in the future, as we learn the art of using phased arrays, and as new technology becomes available.

The sensitivity of a phased array as function of frequency has a simple explanation. The basic element antennas in a phased array ‘station’ (here the term used for the cluster of element antennas) are sensitive to the full sky hemisphere. Each antenna has a typical effective collecting area $A_e \sim \lambda^2/3$, while below 240 MHz the system

temperature is dominated by the brightness temperature of the sky (rather than the noise of the electronics). In this case we have a system temperature $T_{\text{sys}} \sim 60 \lambda^{2.6}$ [with T_{sys} in K and wavelength λ in m], which leads to an almost frequency-independent sensitivity (when expressed in Jansky, $1 \text{ Jy} = 10^{-26} \text{ W m}^{-2} \text{ Hz}^{-1}$) [Bregman, 1998]. A striking consequence is that a station with only 15 simple dipole-like antenna elements is already more sensitive than the 74 MHz mode of a 25 m dish antenna at the Very Large Array radio telescope in New Mexico. This tantalizing prospect, in combination with newly announced digital processing hardware, triggered me to conjecture that a large low-frequency synthesis array could be built within a reasonable budget. Design concepts for such a system were then developed, which would allow 8 simultaneous, independently pointed beams on the sky, each with 4 MHz bandwidth. This would be a truly revolutionary instrument, ideally suited for deep surveys of the cosmos [Bregman, 1999].

Perhaps equally important as these instrumental possibilities was the idea put forward by Jan Noordam that multi-directional self-calibration could offer a way to correct for the ionosphere-induced phase distortions over the beam of each telescope, using bright radio sources in the sky [Noordam, 2000]. Such distortions had severely limited the imaging performance of all previous low-frequency radio telescopes, including the VLA @ 74 MHz. Application of the proposed calibration method requires not only sufficient sensitivity per baseline, but also that the size of the station beam is reasonably matched to the scale size of the distortions in the phase screen. At these low frequencies, the telescopes operate in the sky-noise-dominated regime. Adequate calibration of the time variable phase distortions requires sufficient sensitivity, i.e. sufficient aperture efficiency and bandwidth to solve for the necessary calibration parameters, for every ionospheric coherence time. The antennas of the VLA and other existing telescopes are too small, as are their aperture efficiency and bandwidth at low frequencies, to allow proper self-calibration. This prevents them from achieving high-quality, high-sensitivity continuum images [Cohen, 2007]. However, the results of their pioneering efforts have been gratefully used for the design of LOFAR.

The need for adequate sensitivity to create the proper conditions for self-calibration has guided the system design of the LOFAR radio telescope, including the associated signal and data processing approaches. The telescope implements the design concepts in an innovative way, and is now operational. Arnold van Ardenne at ASTRON has since suggested that I write a dissertation that could serve as a reference document for subsystem engineers and astronomers; a document in which all the elements that constitute a proper synthesis system are set down on paper in a concise way. The emphasis in this dissertation is on the rationale behind a balanced design, recognizing the fundamental limitations in 2-D Fourier imaging, and of self-calibration in the presence of a disturbing ionosphere. It therefore describes the system design of LOFAR. It also describes how the residual phase

errors limit the calibration quality that is achievable in practical 2-D Fourier imaging, and how this will ultimately limit the final sensitivity that can be realized.

The design began in earnest in 1998, in the context of ASTRON's phased-array antenna development for the SKA [Ardenne, 1997], [Ardenne, 1999], [Ardenne, 2000], [Ardenne, 2002]. First, I carried out a study of design concepts for a complete synthesis array, operating at low frequencies. The results were presented at the SKA symposium in Dwingeloo in 1999 [Bregman, 1999] and at the IAU symposium on low frequency radio astronomy in Pune, India, that same year. A complete concept design was presented at the SPIE meeting in Munich in 2000 [Bregman, 2000a], and was followed by a number of articles in scholarly and trade journals with various collaborators [Bregman, 2000b], [Bregman, 2002], [Schaaf, 2003], [Bregman, 2004a]. The latter papers addressed the key aspects of all subsystems for a complete array. This culminated in 2004 with (i) successful operation of the Initial Test Station at the projected site of the LOFAR core in the province of Drenthe in the Netherlands, (ii) my receiving of the Veder Prize 2003 for this work, and (iii) the presentation of a large set of papers at the SKA workshop in Penticton, British Columbia. These papers have been published in a special issue of the journal 'Experimental Astronomy' [Bregman, 2004b], [Cappellen, 2004], [Maat, 2004], [Veen, 2004], [Schaaf, 2004], [Wijnholds, 2004]. This work also served as input for the Preliminary Design Review for LOFAR.

A leading principle throughout the design process and the initial operational phase of LOFAR has been an observation attributed to the late John Baldwin: "Low-frequency observing is all-sky observing". Contamination from bright radio sources in any part of the sky (or the horizon) will affect the imaging performance in the field of interest. I interpreted this remark as a requirement that calibration and imaging procedures should indeed calibrate the brightest sources anywhere in the visible sky. The basis for such a capability is 2-D Fourier imaging with an (almost) planar array, which makes not only clear how the image scale of the imaged sky changes as the Earth rotates, but also defines how the changing beam of a phased array station could be dealt with. The first all-sky image using this principle of combining snapshot images on a fixed sky grid was made by Stefan Wijnholds in 2004 using the Initial Test Station of LOFAR and was presented at the URSI General Assembly in 2005 [Wijnholds, 2005].

In 2005 and 2006, I provided guidance in the detailed design and careful evaluation of prototypes of the LOFAR subsystems. In 2007, it turned out that the available budget for the antenna stations was insufficient for the planned number of stations. The initial station distribution [Bregman, 2005] that would provide adequate U,V-coverage, and hence robust imaging performance, could only be maintained for the core area of the array, although even here smaller stations with a larger beam were necessary. For the high-priority observations of the Epoch-of-Reionization (EoR), which strongly rely on the sensitivity of the core, the survey-speed is only reduced by a factor two. However, the wider station beams require a more complicated

calibration scheme where telescope-based multi-direction self-calibration solutions are combined to reconstruct the ionospheric phase screen over the entire core. At this time, several European countries showed interest to host one or more LOFAR stations, providing baselines up to more than 1200 km, but not yet filling the gap between 80 and 200 km. Quite fortunately, the international funding for these stations allowed full sensitivity low-band arrays and larger high-band tile arrays. Especially the latter have a beam size that is properly matched to the size of phase-screen distortions, provided these occur during quiet ionosphere conditions.

Many documents of a tutorial nature were written in the period 2007-2008 to guide optimization of the array and station configurations of LOFAR. This led to further investigations of the polarization imaging aspects and the calibration limitations by reduced collecting area and extended baselines. It also became clear that, although the standard textbooks give many fascinating details, they fail to provide insight into what really drives the design of a large synthesis telescope. Especially for the hitherto unexplored case of 100 to 300 stations, as are planned for the first phase of the Square Kilometre Array radio telescope (SKA, see www.skatelescope.org). The latter is being considered by the international radio astronomy community, and could even have as many as 1000 to 3000 stations eventually. With such numbers, the density of stations in the central core area becomes so large that the instantaneous aperture plane will be completely sampled. This will allow high quality snapshot imaging with an almost planar array, providing a large FoV with a single 2-D Fourier transform.

The LOFAR array with more than 64 stations and baselines up to 1200 km has a resolution @ 140 MHz comparable to that attainable with the SKA @ 1.4 GHz using baselines up to 120 km. And indeed, full field-of-view imaging with SKA dishes of ~15 m diameter leads to a read-out time and relative spectral resolution of the cross-correlation system that is still a factor two (s)lower than what is required for imaging the large field of view of LOFAR with its stations of order 50 m diameter. This large FoV, expressed in resolution elements, makes LOFAR a true “pathfinder” for the SKA, especially by requiring new calibration and computationally efficient imaging approaches.

In 2009, the LOFAR experience had a considerable impact on the EU-financed SKA design exercise, in the context of the SKADS program. SKADS required estimates for the processing power that SKA would need for correlation and image formation. Extrapolating the results of conventional imaging packages running on a single core PC suggested prohibitive processing requirements, which would absorb most of the available resources for a SKA [Alexander, 2010]. This dramatic result begged for a detailed analysis of the principles and implementations used in the largely conventional imaging software that was planned for LOFAR, and have resulted in the most substantial part of this dissertation: chapter 3 is devoted to efficient processing for wide-field synthesis imaging.

Bundling the collection of papers that formed the basis of the LOFAR design process would be one way of producing a PhD thesis that describes the design studies and its successful realization in practice, with lessons for the future. Instead, I prefer a more ordered description of the design process and its results. This design is based on elements and approaches that are all well proven, but combined in a new -not yet proven- manner that shows the way forward to efficient processing for wide-field calibration and imaging with the next generation of synthesis arrays with many hundreds of stations, part of which are phased-arrays.

Thus, this dissertation offers an overview of all the new aspects that have been suggested and partly implemented in LOFAR, which have not been applied in older synthesis arrays. The thesis is not a “design handbook” but rather a description of the “*proof-design*” that is the LOFAR radio telescope. The text relates the key system issues, and offers elaborations and intermediate conclusions in subsections, citing papers and reports for more detail. Where needed, it provides background for system designers, subsystem designers and all those who plan to use a modern aperture synthesis array, and wish to understand the basic limitations in obtaining high-quality images.

Glossary

ASTRON	Netherlands Institute for Radio Astronomy
antenna	element that produces an output signal from an incident plane wave
ASKAP	Australian Square Kilometre Array Pathfinder
beam-former	device that performs a beam forming operation
beam-forming	operation where signals from a cluster of antennas are added with a complex weight
CMA	Complex Multiply Add operation
dual-pol	adjective to indicate a pair of orthogonal antennas or signal chains
EoR	Epoch of Reionization
FoV	Field of View
FFT	Fast Fourier Transform
Flop	floating point operation
HBA	High Band Array, also adjective for antennas and station arrays
LBA	Low Band Array, also adjective for antennas and station arrays
LOFAR	Low Frequency Array in the Netherlands
MIT	Massachusetts Institute of Technology (USA)
NRL	Naval Research Laboratory (USA)

phased array	cluster of antennas of which the signals are combined by a beam former
psf	point spread function
RFI	Radio Frequency Interference
sinc	$\text{sinc}(x) = \sin(x) / x$
SKA	Square Kilometre Array
SKADS	SKA Design Studies
snapshot	instantaneous image formed by an array
station	1) antenna cluster of which the signal is cross-correlated with other stations 2) location of one or more phased arrays
TEC	Total Electron Content (vertical column density in ionosphere)
TID	Travelling Ionospheric Disturbance
VLA	Very Large Array (near Socorro, New Mexico USA)
WSRT	Westerbork Synthesis Radio Telescope (in The Netherlands)

The birth of radio astronomy dates back to 1933 when the New York Times published Karl Guthe Jansky's 1932 discovery of radio emission from the Milky Way galaxy at a frequency of 20.5 MHz. Using a number of dipoles that formed an array antenna with a total extent of 30 m the directivity at the wavelength of 14.6 m was rather limited but it could nevertheless be shown that the signal was strongest in the direction of the galactic centre. Radio technology developed rapidly and in 1938 Grote Reber detected radio emission from the Milky Way at 160 MHz using a parabolic dish of 9.6 m diameter that he built in 1937 in his own back yard.

In subsequent years most radio astronomical observations have been done using larger dish antennas and shorter wavelength to improve sensitivity and resolution. The scientific requirement for higher angular resolution drove the use of arrays of dishes operating in interferometric modes and making use of Earth rotation in the technique called Earth rotation aperture synthesis.

Aperture synthesis traces its roots to the 1930s when the Van Cittert-Zernike and the Wiener-Khintchine theorems were formulated. These theorems relate the strength of the electromagnetic field across an aperture to the brightness distribution on the sky. That is, the spatio-temporal cross correlation of the field on an aperture is the Fourier transform of the brightness distribution on the sky as a function of frequency [Brouw 1971]. This formalism allows combining independent interferometer measurements into a single image of the sky.

Early aperture synthesis telescopes included especially the low frequency instrument of Martin Ryle at the Radio Astronomy Group in Cambridge (UK) that demonstrated the technique, which earned him a Nobel Prize, together with Antony Hewish in his discovery of pulsars. In the early 1950s a 4 element interferometer was built to survey the sky and produced the 2C catalogue of radio sources at 81.5 MHz and the 3C catalogue at 159 MHz that were both published in 1959. The first production telescope of the type was the Westerbork Synthesis Radio Telescope (WSRT) for studying the 21cm line of hydrogen, which began operation in 1970.

Low frequency astronomy suffers from both the phase distorting effects of the ionosphere and the presence of man-made as well as naturally occurring radio frequency interference (RFI). But scientific interest has been strong enough to lead to several experimental arrays. In the USA Bill Erickson built the Decametric array at Clark Lake, using thin wire half-wave dipoles but his scanning array was T-shaped with an extent of 3 km in EW direction and 1.8 km in NS direction. The first publication dates back to 1965 describing a system working at 26.3 MHz. The largest low frequency radio telescope for many years, having a maximum collecting area of about 150,000 m², was erected in the early 1970s in the Ukraine using 2040 fat

dipole antennas of 9 m length and about 2 m diameter. The antennas of the UTR-2 telescope allow operation at 8 – 40 MHz and are placed in a T-shaped array with 1.8 km extent in NS direction and 0.9 km in EW direction. Erickson subsequently improved the Clark Lake array using 720 conical shaped antennas with a typical length of about 8 m and largest diameter of 4 m that allowed observing over a frequency range of 15 – 125 MHz with best performance between 25 – 75 MHz. A further upgrade combined 15 antennas per bank and the 48 banks were cross-correlated to form an aperture synthesis imaging instrument of which description and first results were published in 1982. The Cambridge Radio Astronomy Group built the Cambridge Low Frequency Synthesis Telescope using 60 tracking Yagi antennas operating along a 4.6 km EW track and published the first part of the 6C survey at 151 MHz in 1985 and the last part in 1993. Another 60 Yagi antennas operating at 38 MHz were added and the 8C survey was published [Rees, 1990].

Since 1998 all 27 dishes of the NRAO Very Large Array in the USA could be used with an additional prime focus antenna that allowed observing at 74 MHz with a limited bandwidth of 1.5 MHz, which provided a large sky survey mainly using baselines up to about 10 km [Cohen, 2007]. The quoted rms noise is 0.1 Jy beam^{-1} typically although the expected thermal noise is only 35 mJy beam^{-1} . However, the point source detection limit of the survey is set at 0.7 Jy beam^{-1} , which is a factor 4 higher than could be expected from a 5-sigma threshold for thermal noise alone. These numbers reflect the problems associated with calibration of ionosphere induced phase disturbances. At low frequencies, these problems can in principle only be overcome when certain array and station design requirements are met [Bregman, 1999], [Noordam, 2000], [Kassim, 2000], [Cotton, 2004], [Lonsdale, 2005], [Thompson, 2006], while this dissertation gives a detailed analysis for the additional to be expected noise contributions.

LOFAR as pathfinder

A major breakthrough in the field came with the realization, that cheap wide band antenna systems are possible when impedance matching between antenna and receiver is of minor concern [Ardenne, 1999]. This is a special situation, which occurs when the noise temperature of the low noise transistor amplifiers is heavily dominated by the sky brightness temperature, as could be the case for frequencies below about 300 MHz [Bregman, 1999]. Instead of large fat dipoles or long log-periodic antennas a simple thin wire short dipole can then be used [Tan, 2000] of 1.4 m height above a ground mesh of $3 \times 3 \text{ m}^2$ [Arts, 2003]. The cost of electronic digitizing and signal processing had by about 2002 come down such that digitizing the full band from 15 - 88 MHz including HF and VHF transmissions would become affordable. With such digital processing all interfering terrestrial emission could in principle be eliminated by spectral and spatial filtering [Ellingson, 2003].

Instrumental in the realization of a new dedicated low frequency instrument was the notion that, following Moore's law, the huge processing capacity needed to correlate such large bandwidth from interestingly large numbers of antennas (for example of order a hundred) and to self-calibrate and image large sky fields at low frequencies, would just become affordable shortly after 2003 [Bregman, 2000a].

Progress in electronic signal reception and processing as well as in data transport over optical fibre still continues and promises a cost decrease that makes large aperture arrays the design of choice for not too high frequencies. The dominance of dish based aperture synthesis arrays is therefore pushed to higher frequencies. For SKA, which was conceived to study neutral hydrogen over most of the history of the universe, phased array technology for the frequency range up to 1.4 GHz could in principle take over by 2020.

In this context, the present dissertation provides a systems oriented background for wide field imaging and the processing aspects of synthesis arrays using phased array stations (stations being the term used to describe clustered antenna arrays). It gives references to details on array and station configuration as well as other system design aspects that are presented in separate papers [Bregman, 2000b], [Bregman, 2002], [Cappellen, 2004], [Bregman, 2004b], [Cappellen, 2006], [Bregman, 2008], [Norden, 2010], [Bregman, 2010], [Wijnholds, 2011].

Elementary antenna properties and basic sky properties such as total brightness temperature and source distribution allow for low frequency arrays where the system temperature is determined by an average sky brightness temperature. For such a low frequency array the single requirement of sky noise limited imaging performance then allows deriving (i) the global requirements for the configuration of the synthesis array and (ii) the size and configuration of its constituting phased array stations that allow wide field image forming [Bregman, 2005], [Wijnholds, 2008]. An implicit assumption is that the distortions caused by the ionosphere are corrected by self-calibration, and efficient processing algorithms properly handle all other unwanted side effects. Minimizing the total processing cost for station beam forming, array correlation and image forming as function of required field-of-view (FoV) provides an additional input for array and station configuration [Bregman, 2004a], [Alexander, 2010], [chapter 6].

Together with technical and scientific boundary conditions, a design has resulted that has been implemented as the LOFAR radio telescope.

LOFAR [Vos, 2009], [Haarlem, 2012], is an aperture synthesis array based on technologies such as (i) short dipole antennas arranged in large phased array stations, (ii) new digital receiver technology that also handles man-made signals in the two bands (10 – 90 MHz and 115 – 240 MHz) shared with active spectrum users. Instrumental is (iii) new optical fibre transceiver technology that makes wide band signal transport up to 80 km cheap and over more than 1200 km affordable [Maat, 2004] using own and shared fibre networks.

These technologies make possible a hierarchical signal processing design [Schaaf, 2003] that allows optimisation for cost [Bregman, 2004a] and imaging performance. The resulting design of LOFAR is a processing dominated system using programmable chips that form the station beams [Gunst, 2005], while new routing technology allows signal transport to a High Performance Computer where tens of thousands of processing nodes perform not only the cross-correlations [Romein, 2006] between antenna stations but also remove RFI and correct the signal streams. Essential new technology is the direction-dependent multi-source self-calibration software that handles the ionospheric wave-front disturbances, removes the strongest sources from the correlated signals, and provides corrections for field distortions [Noordam, 2006], [Tol, 2007], [Intema, 2009], [Smirnov, 2011]. This calibration approach forms the basis for new imaging approaches that treat the varying shape and polarization characteristics of the primary beam of the phased array antenna stations and provide distortion free synthesis images where the noise floor is determined by the global noise of the sky itself.

LOFAR differs from existing low frequency arrays in providing successful imaging in the face of ionospheric phase distortions with (i) up to 50,000 spectral channels, (ii) fields of view with $2^0 - 10^0$ diameter, (ii) details at the 10 arcsec level, and (iii) resolutions down to 0.2 arcsec when the longest European baselines are used at ~ 200 MHz. An even more important characteristic (iv) of the LOFAR synthesis array is the varying station size that allows appropriate direction dependent self-calibration over array scales from 1 km to over 1200 km to correct for ionosphere phase disturbances. Instrumental in this respect is a useful instantaneous bandwidth up to 90 MHz allowing relative bandwidths of $\sim 20\%$ for multi-beam continuum observing. Such bandwidth gives not only sufficient sensitivity for multi-source self-calibration but provides for limited gaps in the U,V-coverage (i.e. Fourier domain sampling) of a long synthesis observation with a limited set of stations. Good U,V-coverage is essential in reducing artefacts in synthesis images that can easily cause image noise far stronger than potentially possible with the investment in collection area and processing bandwidth.

This dissertation presents the approaches as used during the design of LOFAR as a wide-field polarimetric imaging telescope at low frequencies. Detailed design progressed in the period 2000-2005 where it passed a Preliminary Design Review and a Critical Design Review. A major change in shape of the LOFAR array was realized in 2007 when the European stations had to be included within a reduced budget. In the period till 2011 the scientific base under the various design approximations has been strengthened, so a generic approach is described that can be scaled and is applicable to arrays with phased array stations operating below ~ 0.5 GHz, which means that LOFAR is a true pathfinder for the low frequency segment of the SKA.

Structure of the dissertation

There are two distinct parts to the work.

The overview, in chapter 2, shows all the system aspects that make LOFAR different from existing arrays and which allowed detailed design of an affordable low-frequency synthesis array using phased array antenna stations, while more detail and analysis can be found in the references. It will be shown how all these elements relate together and need an appropriate configuration of antennas in stations, and of stations in a synthesis array, to provide wide field images in which the sensitivity is limited only by thermal noise, thereby minimizing the processing requirements and therefore cost.

The second, main part of the work is a more detailed imaging and calibration analysis that derives requirements for the array and station configurations as well as for the calibration and imaging procedures. The focus is on the fundamental limitations involved and on potential implementations that minimize the total processing effort, which is essential for future larger instruments. Since the actual LOFAR configuration satisfies these requirements only partially, we expect that in due time high quality high-resolution imaging performance could be reached only over parts of the full field of view provided by the largest telescope beams. The most important high quality criterion is reaching sensitivity levels defined by (i) total collecting area, (ii) thermal noise of sky and receiver (iii) bandwidth of the instrument, and (iv) total integration time, which can cover many weeks.

The core of the dissertation is formed by analyses of additional image noise by residual phase errors over the field of view of a station beam introduced by phase errors in self-calibration and approximate 2-D Fourier imaging as function of configuration. It is the propagation of these phase errors to relative errors in the nominal side lobes of the point spread function (psf), which will ultimately limit the sensitivity of continuum images, when the nominal side lobes are effectively subtracted. This effect is the main driver for array configurations to provide full U,V-coverage such that the psf side lobes are low as determined by the applied taper function, which also reduces the associated errors in these side lobes that are caused by residual complex gain errors in each station beam. When the same field is for instance observed a hundred times, thermal noise and ionosphere induced phase errors reduce by a factor ten. However, systematic effects, as could occur in imaging, are not averaged out and could even start to dominate the noise floor in the integrated image.

In chapter 3 we analyse the fundamental limitations in Earth rotation synthesis imaging and derive expressions for the phase effects that limit the accuracy of wide-field imaging approaches as function of non-planarity of the baselines in a planar Fourier image. We address the specific effects of phased array stations, such as effective position, foreshortening, polarization, and so-called blind angles, and de-

rive a characteristic time interval for correction thereof. We present two new methods of processing efficient imaging and compare the minimum processing requirements with those for two existing methods. The most important conclusion is that the number of sources, which have to be subtracted to reach the thermal noise, dominates the processing of the most efficient methods for continuum imaging.

Chapter 4 summarizes the multi-source self-calibration approach adopted for LOFAR and analyses theoretical and observational results to derive the characteristic time scales and scale sizes of ionosphere induced wave-front distortions. The approach uses a limited number of strong sources per station beam that allow proper self-calibration and accurate subtraction of these sources. The derived calibration parameters per station can then be used to find interpolated corrections for all other sources in the field allowing subtraction of these sources with limited accuracy.

We evaluate published differential source counts data to derive an integrated source count formula for the frequency range of LOFAR. Given the sensitivity of the LOFAR interferometers, we estimate the expected number of sources per station beam that have a signal to noise ratio (SNR) larger than 3 on a sufficient number of baselines per station to provide self-calibration of each station. It turns out that this number of sources is for LOFAR sufficiently large to provide interpolated corrections over a full station beam using integration times comparable to the derived ionosphere coherence time. It will be shown how observed interferometer phase as function of frequency can be used to estimate the *variation* in total electron content (TEC) over each station beam without running into potential ambiguity problems. Refraction as function of elevation over the station beam would in principle allow estimation of *absolute* TEC that can be used for correction of Faraday rotation using a model of the Earth magnetic field.

Apart from mathematical interpolation errors, which can be made sufficiently small if large scale phase gradients are sampled sufficiently dense, we also suffer from physical interpolation errors, since large-scale phase gradients, dissipate to smaller short-scale ones by Kolmogorov evolution. These two effects define a minimum phase error for each direction in each station beam as function of distance from the reference directions.

Chapter 5 analyses the contribution to the image noise by the nominal side lobes of sources in a field, as well as the contribution by error side lobes due to phase and amplitude errors over the station beam. An important property of this error propagation is that small phase and amplitude errors per station cause an error pattern in the psf that is proportional to the square root of the nominal psf of a snapshot image. This shows the importance of good U,V-coverage per snapshot providing low nominal psf side lobes in the first place that could even be lowered further by appropriate tapering if almost full snapshot U,V-coverage is provided.

We give a first order estimate of the nominal side lobe distribution of a narrowband snapshot psf, and show how that distribution evolves as function of increasing bandwidth and tracking time. This nominal 2-D psf determines the total number of sources that have to be subtracted from the projected 3-D visibility data of each instantaneous snapshot dataset to make the thermal noise dominating over the noise by the sum of the remaining side lobes of all sources that are not subtracted. In addition to this nominal side lobe noise there is a noise contribution by errors in the nominal side lobe pattern of sources in between the self-calibration sources, due to phase errors over the station beam. This image noise contribution cannot be removed by subtracting more sources, but needs a denser grid of self-calibration sources to be used for interpolation.

Chapter 6 presents scaling laws for the processing required by correlation and continuum imaging as function of system configuration and summarizes the key aspects that allow high imaging quality at affordable processing cost and gives recommendations for LOFAR as well as for larger arrays such as the low-frequency segment of the SKA.

2

Overview of System Design for LOFAR

LOFAR is designed as a versatile aperture synthesis array for many science applications [Kassim, 2000] that in the first place will be used as imaging instrument. Although its sensitivity as defined by the total collecting area is only slightly larger than that of the UTR-2 in operation in the Ukraine since the early 1970s, the total collecting area is distributed to form a high resolution aperture synthesis array that is no longer confusion limited.

LOFAR [Vos, 2009], [Haarlem, 2012], has two sets of antennas that cover a much larger frequency range than earlier instruments, spanning 10 – 90 MHz and 115 – 240 MHz respectively. Another important difference is its digital signal processing capability that can handle the strong signals of regular transmission in instantaneous bands of up to 100 MHz and recover the weak astronomical and also terrestrial signals with spectral resolution narrower than 1 kHz and temporal resolution down to 5 ns. Finally, the total collection area of LOFAR is distributed over more than 64 phased array stations of which the signals can be cross-correlated. The baselines span a range from 120 m to 1200 km providing spatial resolution up to 0.2 arcsec at its highest frequency.

LOFAR provides not only a sharper eye on an up till now poorly explored part of the observable electromagnetic spectrum of our Universe, but the new digital receiver and signal processing capabilities will extend the number of pulsar detections and allow imaging of the evolution of Jovian lightning flashes. In addition, terrestrial lightning flashes and even the radio flashes that accompany the particle cascade that could emerge after penetration of very high energy neutrinos in the Earth atmosphere could be imaged in 3-D space [Falcke, 2006]. Although LOFAR opens up a new window for a broad range of scientific research, one of the most important key science projects is to detect the signal of the Epoch of Reionization (EoR) with the core stations of LOFAR [Rötgering, 2006].

LOFAR has, many applications but our focus is on wide field continuum imaging. From the design point of view, this is one of the most demanding applications, especially when fields are observed many times to reduce the thermal noise reaching a level where systematic effects due to calibration and imaging could start dominating the noise floor. This makes clear that calibration and imaging procedures form an integral part of a system design in which the configuration of the synthesis array and the configuration of the antenna elements in the phased array stations form the basic ingredients. However, these basic ingredients need to be designed such as to support optimum observing, calibration and imaging procedures that have their own limitations and pose requirements on array and station configuration. In separate

papers details of the LOFAR array configuration [Bregman, 2005], [Bregman, 2012] and the station configuration [Wijnholds, 2008] are presented.

The chapter is organized as follows.

The first section of this overview chapter describes the global design drivers of LOFAR.

The second section summarizes the main characteristics of LOFAR to show the features of an aperture synthesis array where the individual antennas are clustered into phased array stations.

The third section summarizes calibration and imaging limitations as encountered in low frequency observing that limited the performance of previous synthesis arrays.

The fourth section here addresses the processing issues for high resolution wide field low-frequency imaging and shows that real time imaging can in principle be done by a post correlation processing platform with less power than the platform used for cross-correlation, if dedicated calibration and imaging procedures are implemented.

The fifth section on new approaches in the design of LOFAR presents all the key items that explain why LOFAR could be designed and realized and why we expect that LOFAR will succeed in making high quality high-resolution low frequency images.

2.1 Global design drivers for LOFAR

Detection of the very weak EoR signal with LOFAR is only the first step towards imaging of large-scale structures of this transition, which is in principle possible with the low frequency segment of a future SKA. It is therefore essential that LOFAR demonstrates that it reaches its nominal imaging sensitivity in a 6 h synthesis observation, and that this level can indeed be improved by a factor 10 when 100 of such observations are averaged. Therefore, the focus of the design of LOFAR has been on imaging performance in aperture synthesis mode, which is also the focus for this thesis.

2.1.1 Design for Imaging

Imaging is based on the Van Cittert-Zernike theorem where a 2-D Fourier transform relates the cross-correlation function of the electric field distribution in an aperture plane to a distribution of remote objects radiating incoherently. The theorem is

based on a number of assumptions [Thompson, 2004] such as complete sampling of the aperture plane, for instance with a set of interferometers. In practice less than complete sampling is the rule and a Fourier image then suffers from artefacts that can be removed effectively if certain additional image constraints are fulfilled. An important sky property is that the solid-angle density of isolated source structures stronger than a certain flux density threshold increases monotonically if the threshold is lowered. This means that for a given sensitivity limit only a limited number of sources will be present in a Fourier image. In such a case it can be argued that a limited number of independent interferometer observations covering a limited field of view will be adequate to describe the image fully, except for structures buried in the noise.

A given contiguous collecting area A_c with some appropriate tapering could provide a beam on the sky with a solid angle Ω_c approximately given by $1.5 \lambda^2 / A_c$, where λ is the wavelength of the observation. If there are a number N_{sc} detectable sources, within the solid angle Ω_c , they cannot be separated and detected individually. When the total antenna collecting area is separated into N_{st} smaller stations $\frac{1}{2} N_{st} (N_{st} - 1)$ interferometers can be formed and a total image aperture area could be sampled that is larger than the antenna aperture by a factor $\frac{1}{2} (N_{st} - 1)$. This image aperture need not be contiguous and could then provide larger spatial frequencies. The solid angle of a single station beam is a factor N_{st} larger than Ω_c and therefore the total number of sources observed by the synthesis array that exceed the detection threshold is also increased by this factor N_{st} . All isolated sources in a field defined by the beam of a station can now be imaged in principle if we satisfy $\frac{1}{2} N_{st} (N_{st} - 1) > N_{st} N_{sc}$, which explains why aperture synthesis works in the first place.

A more detailed analysis needs to include the increase of the number of independent interferometer samples by Earth rotation when the total observing time is extended. For continuum observations, the sampling of spatial frequencies in the aperture plane can be made denser by increasing the spectral range, but only if it can be assumed that source structure is independent of frequency, as discussed by [Rao, 2010]. Especially if the sensitivity is increased by repetition of identical observations in an attempt to detect more continuum objects of lower intensity we need to realize that no additional information about the source distribution is observed and that we need to design the array with sufficient stations for that situation.

Sparse sampling of the aperture plane has led to successful imaging at frequencies where the sky is dominated by small numbers of isolated sources. At low frequencies, the large FoV defined by the station beam contains large numbers of isolated sources requiring sufficiently long baselines to prevent confusion. However, the large-scale emission requires sampling at very short baselines as well. An appropriate sampling of the Fourier plane is then a very hierarchical sampling, even one approaching a fractal distribution.

2.1.2 Sensitivity types and impact on instrument design

An aperture synthesis array is basically characterized by three sensitivities: Sensitivity to Point sources, sensitivity to Surface Brightness, and Survey sensitivity.

The total collecting area is the prime system parameter that determines the sensitivity for detection of unresolved sources, and is the main cost driver.

The distribution of the stations outside the core determines the fraction of U,V-samples that contributes to the surface brightness sensitivity of partly resolved objects. The requirement to have reasonable brightness sensitivity over a large range of resolutions then drives to an exponential distribution of stations [Thompson, 2004].

The sensitivity that can be reached in a given duration of a survey that covers some fraction of the sky depends on the total FoV of a synthesis array. If this FoV is smaller than the survey area, the sensitivity per survey area can be increased by extending the FoV of the instrument. Such an extension could for instance be realized by forming more beams using the same station aperture and means replication of cross-correlation, calibration and image processing of these additional beams. This form of extension increases the total cost, but has no impact on array or station configuration. However, the maximum number of digital station beams determines the minimum number of antennas in the focal plane of a dish or the minimum number of antenna clusters that together provide a digital signal in a phased array station.

Finally we need sufficient sensitivity per baseline to allow self-calibration that corrects for time-varying ionosphere induced phase deviations in the received source wavefronts, as illustrated in sub section 2.3.1 and further discussed in chapter 4.

2.1.3 Minimum station size and calibratability

A dominant source of image distortion and image noise at these low frequencies is the Earth's ionosphere. The ionosphere induces phase variations over the source wavefronts that have a certain angular structure and vary with a characteristic time-scale. In a synthesis array, therefore, not only does one need sufficient numbers of stations that are correctly distributed to provide adequate U,V-coverage for imaging, but the individual stations need to be appropriately sized to provide a beam matched to these angular phase structures. Also the station needs sufficient sensitivity to be able to detect in roughly 10 seconds of integration adequate numbers of sources across the sky to generate an instantaneous map of the induced phase distortions, as will be discussed in chapter 4.

The system temperature for well-designed receiver systems, at least for frequencies below 240 MHz, is defined by the sky brightness temperature, $T_{\text{sky}} \sim 60 \lambda^{2.6}$ [with

T_{sys} in K and wavelength λ in m]. Phased array stations working in the sparse regime have an effective aperture proportional to wavelength squared [Cappellen, 2004], which leads to a System Equivalent Flux Density per station that increases with $\lambda^{0.6}$ (when expressed in Jansky, $1 \text{ Jy} = 10^{-26} \text{ W m}^{-2} \text{ Hz}^{-1}$). Including a bandwidth that is proportional to frequency we get a detection sensitivity almost proportional to wavelength. The flux density of most calibration sources increases with wavelength almost at the same rate and the number of detectable sources per unit solid angle is therefore almost independent of frequency.

However, the solid angle of the station beam is proportional to wavelength squared and leads to less detectable sources per station beam at higher frequencies. Therefore, we need to increase the number of antenna elements per unit station aperture at higher frequencies to provide sufficient self-calibration sources per beam per ionosphere coherence time [Wijnholds, 2011]. LOFAR uses this principle of varying antenna separation in the LBA stations, which will be further discussed in subsection 2.2.2.

A relative bandwidth of ~20% is sufficient for to provide in an ionosphere coherence time sufficient sources suitable for self-calibration. With 5 sources a 2nd order interpolation scheme allows estimation of phases over the whole station beam. Such an interpolation is sufficiently accurate if the ionosphere is appropriately sampled by the 5 sources in the beam. The dominating medium scale travelling ionospheric disturbances (TID) then need a beam size less than 4° [Thompson, 2006], [Wijnholds, 2011]. This requirement drives to large stations at low frequencies, while the number of stations should also be sufficiently large to solve for a multi-parameter solution and provide sufficient image quality for instantaneous imaging.

2.1.4 Global Design Considerations

LOFAR is the first aperture synthesis array where the stations are phased arrays, which means that the station beam not only rotates but also changes its shape with respect to the sky when a field in the sky is tracked while the Earth rotates. This is a new aspect in aperture synthesis that needs to be handled by appropriate imaging procedures, which make them part of the total design effort. Our focus in this section is on these and other issues that define the new paradigms that make LOFAR different from pre-existing arrays and especially how these issues have driven the final design.

The arguments about minimum station size and minimum number of stations set out above make clear that a low-frequency aperture synthesis array needs a minimum total number of antenna elements to reach sky noise limited imaging performance [Bregman, 1998]. Consequently, a minimum budget is needed to provide images with a quality standard set by observations at higher frequencies.

At low frequencies, sparse phased arrays are the design of choice that provides maximum sensitivity over more than an octave bandwidth within a given budget [Bregman, 2000a]. An important design aspect is the possibility of flexible distribution of the total number of antennas and receiver chains over small and large stations to satisfy the requirements of calibratability and brightness sensitivity, leading to an exponential shell configuration of stations within the synthesis array [Bregman, 2005].

Realizing that LOFAR is an intermediate step towards the SKA [Ardenne, 1999], [Ardenne, 2002], it could be argued from the design point of view that LOFAR is then an ideal test bed to implement promising technologies and new approaches that will qualify as proven technology by the time that the SKA has to be designed and materializes. This has led to implementation of a design based on elements of not always fully proven technology for the specific LOFAR application and circumstances, but fully justified to address the scientific challenges that bear their own risks.

Cost control benefits from proper balancing between (i) the cost for signal collection, which is driven by the total number of antennas and receivers, (ii) the cost for signal transport, which is driven by the number of stations, and the distance to the furthest stations, and (iii) the cost for further processing, which is driven by the longest baselines. High resolution in wide field continuum imaging drives the processing cost that is proportional to total FoV expressed in resolution elements. Sufficiently long baselines are however essential to bring the source confusion 'noise' (caused by unresolved sources) to a level below the thermal sensitivity limit provided by collecting area, system temperature, bandwidth, and integration time.

A golden rule in system design is that an optimum performance over cost ratio is reached when the marginal performance-over-cost-ratio of all main constituents are all equal. Of course, most cost effective technology is assumed but complications arise in defining appropriate performance metric and appropriate boundaries for the main constituents [Bregman, 2004a]. Also the cost metric needs to be defined carefully and total cost of ownership over the expected lifetime of the system is then most relevant. Especially for systems where non-recurrent engineering cost is not only dominant but is also financed separately from the system realization budget, a non-optimum system might result. It could sometimes be argued in such a case that at a higher systems level still an optimum allocation of resources is obtained to realize performance goals at that higher level.

An instrument designed for maximum survey sensitivity could use up to 50% of its total cost in receiver electronics and platforms for signal and data processing [Bregman, 2004a]. The actual design of LOFAR realized computational robustness and adequate sensitivity within a limited budget and is indeed found to be a processing dominated system [Schaaf, 2004]. In this regard, we need not only consider the transformation of digitized antenna signals at the stations into a set of sub

bands from which digital beams are formed [Gunst, 2005], but also the cross-correlation of the station signals including the RFI flagging of the correlation output and binning in appropriate channels with appropriate integration time [Romein, 2006], [Vos, 2009]. These arguments point to an equal distribution of processing capacity at station and array level [Bregman, 2004a] and optimally to a number of dual-polarization receiver chains per station equal to the number of stations, which is indeed approximately true for LOFAR.

Further processing steps involve the direction-dependent multi-source self-calibration [Tol, 2007] including the subtraction of the few hundred strongest sources [Nijboer, 2006], the creation of snapshot images, corrections for rotating beam and polarization and the combination into a wide field synthesis image. Finally, we need deconvolving the strongest side lobes of all remaining sources, which would otherwise determine the noise floor in the final images.

2.1.5 Processing cost evolution over time

We conclude this overview of design considerations with the realization that our new approaches have all been made possible by the evolution of 21st century electronics that make large scale transport of digital data and processing thereof affordable. By relying on Moore's law we could already start designing a low frequency array based on phased array stations in 2001 while it would only be affordable if all processing elements would be ordered after 2003 [Bregman, 2000a]. This allowed us to develop in due time implementations that overcame the conventional limitations in array processing.

Since price erosion of digital signal transport and processing equipment is still expected to continue until 2020 this begs for a staged approach in the realization of the SKA in steps. When every 3 year the total collecting area is increased by a factor of order 3 by adding more stations, also a new correlation platform is required that has 9 times the processing power of the previous one, but only three times the input bandwidth. A detailed analysis then shows that the cost increases by only a factor 3 keeping the cost constant as fraction of the total investment [Bregman, 2010]. This is contrary to the approach followed, for example, by the Atacama Large Mm Array. A correlation platform has been designed around a dedicated custom integrated circuits developed in the late 1990s. Although a cost optimum then the system could not profit from technology advances during its prolonged installation.

LOFAR is the first array of which the design is based, not on fully proven technology, but on preliminary performance specification of commercial electronic components and processing platforms that can be easily upgraded. This pathfinder approach shows the path towards a cost effective low-frequency segment of the SKA with a field of view of hundreds of square degrees provided by phased array stations that operate up to frequencies of ~ 0.5 GHz.

2.2 LOFAR Characteristics

LOFAR [Vos, 2009], [Haarlem, 2012] is a synthesis array centred in The Netherlands at 53° North and 7° East. It has extensions in other European countries with coordinates ranging from 60° North to 45° South and from 5° West to 22° East, and even to 35° East when an Ukrainian station is included. LOFAR operates in two frequency bands, with two sets of antenna arrays. The Low Band Array (LBA) covers 10 – 90 MHz while the antennas are sky noise limited between 30 and 80 MHz. The High Band Array (HBA) covers 115 - 240 MHz, but has about 50% aperture efficiency at 190 MHz and lower at higher frequencies. The gap between 90 and 115 MHz ensures minimal sensitivity to the commercial FM radio bands which are very strong across Europe.

The synthesis array is formed by antenna stations that are phased arrays themselves, so stations and array share the property that the angular resolution in elevation is to first order inversely proportional to the sine of the elevation angle. The sensitivity of a station is limited by the characteristics of the beam of its element antennas and decreases rapidly at low elevation.

2.2.1 Array

When realized, the array is configured for ~40 stations each with 48 dual polarization receiver chains in The Netherlands and has 8 stations with 96 receiver chains in other European countries. About 24 of the Dutch stations are placed in the central core area near the village of Exloo, and the remote stations are placed at distances of up to ~80 km.

The configuration of the core and close by remote stations has been optimized for U,V-coverage after 12 h observing at high declination. The location of the remote stations has initially been chosen [Bregman, 2005] with 5 spiral arms to give good U,V-coverage after ~5 h using a relative bandwidth of ~20%, which allows for multi-frequency synthesis [Rao, 2010] and gives sufficient sensitivity for self-calibration. The shorter time interval is important to avoid low elevations for sources with low declination since phased array stations have low sensitivity at low elevation while ionosphere induced phase disturbances are strong.

The European stations go out to a radial distance of ~600 km and in future possibly as far as ~2000 km. Their placement has been opportunistic rather than driven by any optimization algorithm. Future enhancement with some stations at distances between 80 km and 200 km from the core might improve the baseline distribution for high quality imaging at the highest resolution (actual data can be found on the LOFAR section of the ASTRON website).

2.2.2 Stations

The Dutch LOFAR stations have a varying size and contain two or three sub arrays. All stations have a Low Band Array (LBA) with a diameter of ~81 m that has 96 dual polarization antennas placed in expanding shells.

The remote stations have a High Band Array (HBA) with a diameter of ~40 m consisting of 48 tiles of $5 \times 5 \text{ m}^2$, while the 24 core stations each have two small HBAs with a diameter of effectively ~28 m that are ~130 m apart and each has 24 tiles. A tile is a structure for mounting element antennas for ease of handling and protecting. Every tile has 16 dual polarization antennas of which the signals are combined by a true time delay beam-forming network for each set of linear polarized receptors.

Each digital receiver chain is connected to one HBA tile and to two LBA antennas and allows selecting from the LBA those 96 single polarization dipole signals for digital beam forming that allow optimization of beam width and maximum effective aperture for a limited frequency range [Nijboer, 2009].

The elliptical core of the array with axes of 1.9 km and 2.4 km has 46 small HBAs and 2 more HBAs at short distance of which the signals can be cross-correlated. This gives an almost circular core beam for most of the fields that pass the array at 40° elevation at meridian transit.

The 8 European stations have 96 tiles providing a HBA of 56 m diameter with a narrower beam for better matching to local ionosphere patch sizes, while the LBA has a diameter of 68 m and has an antenna distribution that is better optimized for observing above 50 MHz.

2.2.3 Low Frequency issues and interference

The LBA stations vary in effective size from 32 - 81 m in diameter, which is pretty large compared with standard parabolic dish antennas, but has become affordable thanks to new technology. However, when measured in wavelength (ranging from 3 – 20 m) the station size is still limited. The consequence is that the station main beam is wide and that the side lobe level is high. This observation has led John Baldwin to his famous saying that low frequency imaging is in fact all sky imaging, and I would add that observing should better be organized that way. Indeed many key science projects are surveys, but calibration and image forming is still organized with focus on the sky area covered by the main beam. However, the troublesome part in receiver design is in handling the strong man-made signals in the station side lobes that can produce spurious signals.

Regular transmitters that operate in their allocated frequency bands, where LOFAR observes as well, form an important class of troublesome signals. These sources can just be handled as regular strong sky sources, for which a proper calibration will be made for every instant. The objects can then be subtracted, or in more appropriate signal processing terms, their signal can be projected out [Ellingson, 2003], [Wijnholds, 2004], without disturbing the data for further image forming. It has been demonstrated [Boonstra, 2005] that even interference created in the receivers by cross modulation of external signals can be handled this way.

2.2.4 Signal processing at station and array level

Current technologies provide extraordinary flexibility in the signal processing capabilities for LOFAR.

The 96 digital receivers (192 at the remote European stations) each produce 512 sub bands of 195 kHz covering 100 MHz bands that are limited to provide effectively ~80 MHz bandwidth. There are for each polarization 512 digital beam formers that can be controlled to provide 512 independent beams on the sky where each so-called beamlet selects one sub-band from each of the 96 (single polarization) antenna inputs [Gunst, 2005]. The beamlets can be configured such as to provide a single station beam with bandwidth up to 100 MHz of which smaller fractions are effectively passed by filters in the receiving systems, or more beams in different directions each with smaller bandwidth in single or dual polarization.

The stations provide also a full set of polarized cross-correlations between all elements for a selectable sub-band. When an integration time of 1 second is chosen the full bandwidth is after 9 minutes available for station calibration. With ~40 stations that step differently through their sub bands a total instantaneous bandwidth of 8 MHz is available every second for all sky monitoring, for instance for solar bursts, lightning strokes, etc.

The array correlation system implemented on the Blue Gene/P Supercomputer located in Groningen processes 3.1 Gbit/s per station, which converts to 48 MHz bandwidth with complex samples of 2 x 16 bits in two polarizations for one beam. Alternatively 8 bit and 4 bit modes are available providing more bandwidth and more beams respectively, and for special modes the core stations alone could provide even larger data bandwidth to the central correlation system.

2.2.5 Field-of-View

Hierarchical clustering of the element antennas also provides extraordinary flexibility with respect to the effective FoV. Large FoV is the hallmark of the phased array stations, and the digital processing system allows generation of multiple station beams simultaneously on the sky. The small HBAs at the stations in the core of LOFAR have at 150 MHz beams with a solid angle of $\sim 16 \text{ deg}^2$ while the HBA at the remote stations cover $\sim 8 \text{ deg}^2$ and the European stations even $\sim 4 \text{ deg}^2$, assuming some taper that marginally reduces the sensitivity. The Low Band Arrays at all stations have a minimum FoV of $\sim 32 \text{ deg}^2$ at a frequency of 40 MHz, which can be adapted for higher frequencies by selecting from the configuration with expanding antenna separations the appropriate subset of elements. Efficient sky surveying requires a total FoV of typically 200 deg^2 that can easily be provided by the multi-beam property of the station beam forming system.

2.3 Calibration & imaging limitations at low frequencies

Calibration procedures as developed between 1992 and 2004 for the VLA Low-Frequency Sky Survey (VLSS) [Cohen, 2007] are reviewed and it will be shown how limited bandwidth and low telescope aperture efficiency prohibited proper self-calibration to reach the potential resolution and sensitivity of the VLA. This issue will be further introduced in subsections 2.3.1 and 2.3.2.

The VLSS observed at 74 MHz has a wavelength of 4 m, which results in a very wide beam of order 10 degrees diameter with the 25 m dishes of the VLA. Conventional synthesis imaging using a 2-D Fourier transform for projected baselines has only a limited distortion-free FoV that is much narrower than this telescope beam. A new imaging method was therefore developed for the VLSS, called polyhedron imaging, where the total beam area is imaged using a large number of smaller facet images.

Application of this method to LOFAR will be computationally prohibitive since more than 950 facets would be needed for imaging a full station beam at 50 MHz with baselines of 90 km, reason to look for alternative solutions that will be addressed in subsections 2.3.3 and 2.3.4.

Subsection 2.3.5 and 2.3.6 introduce the issues related to polarization in phased array stations and the effects of foreshortening by such stations, respectively.

2.3.1 Sensitivity limits calibratability

At low frequencies, the ionosphere is the dominant source of phase disturbances in the wavefronts of the signals from celestial sources. At 330 MHz we observe phase fluctuations over 3 km baselines [Spoelstra, 1996] that reach typically half a radian. These phase fluctuations can be attributed to Travelling Ionosphere Disturbances (TID) in the ionosphere. The most relevant medium scale TIDs create a density fluctuation in the total electron content at a height of 200 - 250 km, and have a wave like structure with wavelength of 100 - 200 km and quasi-periods of 10 - 20 min [Thompson, 2004]. The result is phase variation proportional to wavelength [Kasim, 1993] resulting in a full phase turnover on baselines of order 10 km at a frequency of 74 MHz and even two turns at 38 MHz within timescales of order ten minutes. Apart from these well-behaved structures with a well-defined observable angular size for the TID, observations suffer from turbulence effects [Tol, 2009] with a characteristic scale size of the “seeing cell” that is defined by the area over which the phase variance is 1 rad^2 . The maximum phase disturbance by the TID is proportional to wavelength, while the diameter of the seeing cells is proportional to frequency.

The VLA low frequency system operating at 74 MHz provides an excellent case study of what is required for high resolution, high fidelity imaging at low frequencies. This system has 1.5 MHz bandwidth and $\sim 15\%$ aperture efficiency, which produces in a typical ionosphere coherence time of 30 s limited sensitivity. Only a few fields in the VLSS have one or more sources in the beam that are strong enough to provide with the available sensitivity a signal to noise ratio (SNR) larger than ~ 2 per polarization baseline to allow self-calibration.

The sensitivity in a snapshot image that combines all baselines is a factor ~ 20 higher and allows for every field in the survey to observe ~ 7 sources with $\text{SNR} > 10$. The actual positions of these sources can be compared with their nominal positions from a catalogue valid for higher frequency, and then be used to correct for distortions over the field [Cotton, 2004]. This approach produces images with better overall quality, than images using self-calibration based on only one source. Self-calibration eliminates only the artefacts of the self-calibration source and reduces artefacts of other sources in its near environment, but increases distortions in sources at larger angular distances [Cohen, 2007].

Clearly, image fidelity can in principle be dramatically improved when three or more sources in the beam have sufficient SNR [Noordam, 2000]. Successful demonstration of such an approach using observed data [Intema, 2009] took however a long development time, but promises success for LOFAR that satisfies all further criteria [Wijnholds, 2011]. This will be discussed further in chapter 4.

2.3.2 Image and source distortion relate to station and array size

Source blur will occur in snapshot images when the wave front phase over all array stations is curved and even speckled images result if the wavefront distortion is stronger and irregular. Removal of such blur requires additional sensitivity per station, which is not available with the VLA system. More serious is the fact that with very few calibrators across the FoV, one obtains high quality images only with an array that has an extent that is smaller than the scale size of the wave front disturbances [Lonsdale, 2005]. This condition forced low frequency observations with the VLA to use the B configuration with longest baselines of 10 km, or the BnA configuration with only one arm of 21 km [Cohen, 2007]. The result is significantly reduced angular resolution and increased confusion noise over what might be obtained with stations that are more sensitive.

Synthesis arrays that are much larger than the scale size of the wave front disturbances need for each station a proper phase correction for a number of directions. The LOFAR design handles this issue by using stations that have sufficient sensitivity [see chapter 4] to solve for direction dependent gain and phase for the five strongest source directions for every station, which allows accurate subtraction of these sources.

Moreover the station beams will be narrow enough such that a simple wave front curvature model using the five solutions is indeed adequate to derive corrections for all sources in the field that are detected well enough to subtract the next set of strongest sources with sufficient accuracy [see chapter 4]. If subtraction would use inaccurate complex gain factors we are left with an error pattern in every snapshot image that is related to the true point spread function (psf) of each subtracted source [see chapter 5]. Unfortunately, these residual error patterns cannot be fully de-convolved in a later processing stage and residual side lobes could ultimately determine the effective noise floor of a synthesis image [see chapter 5].

Matching station beam size to the size of the TID induced structures requires stations that are large enough to provide a beam narrower than $\sim 4^\circ$ [Thompson, 2006], [Wijnholds, 2011] and will be further discussed in chapter 4. Larger stations could increase the accuracy of the model that uses a curved phase screen, which is attractive to reduce the influence of sources outside the main beam.

Unfortunately, within a given budget, larger stations allow only fewer of them, which limits the U,V-coverage in a synthesis image and drives up the side lobe level of the synthesized array beam. Such higher synthesized beam side lobes require additional processing power since more sources have to be subtracted to reach the same noise level in each snapshot. The final noise floor will ultimately be determined by three components, (i) the thermal noise, and (ii) the residual side lobes due to calibration errors in the subtracted strong sources and (iii) the nominal side lobes of all weaker sources that are too weak to be de-convolved. These contribu-

tions will be compared in chapter 5 and could then lead to additional requirements for the U,V-coverage.

Spectral line images need in general no separate source subtraction since the process that subtracts the continuum contributions already removed most strong sources. As a result, they will reach a noise floor as determined by the thermal noise as follows from effective collecting area and time-bandwidth product.

2.3.3 Array planarity, Field-of-View and faceted imaging

In this subsection we address the FoV limitations by conventional synthesis imaging and the impracticality of the of the polyhedron method use by LOFAR.

The starting point for synthesis imaging is that according to the Van Cittert-Zernike theorem the coherences measured in the U,V-plane of a planar correlation array can provide the superposition of two hemispheric sky images by simple Fourier inversion [Thompson, 2004]. Deviations from array planarity cause errors in the nominal side lobes of an array beam when the computationally efficient 2-D Fourier beam-forming is used, where the baseline vectors are projected on the reference plane of the transform. In addition, since processing capacity is a limiting resource within the LOFAR system there is a strong drive to use processing algorithms with logarithmic characteristics such as the Fast Fourier Transform (FFT) that can however not include position dependent phase corrections.

An interesting approach is found possible if, instead of looking at the errors in the side lobe pattern, we analyse the phase deviation in the visibilities for a station that deviates a distance H from the reference plane of an almost planar array. At wavelength λ the maximum visibility phase deviation $\Delta\phi_m$ of a source at a nominal and small angle α from the normal on the reference plane is given by [Taylor, 1999] and will be further analyzed in subsection 3.1.8. We get

$$\Delta\phi_m \sim \pi \alpha^2 H/\lambda \quad [\text{rad}] \quad (2.1)$$

If we tolerate a maximum phase deviation of π^{-1} we could define a maximum radius α_m of the FoV of a 2-D Fourier image given by

$$\alpha_m \sim \pi^{-1} (\lambda/H)^{1/2} \quad [\text{rad}] \quad (2.2)$$

For synthesis imaging we conventionally use a reference plane for FFT imaging that is perpendicular to the direction of the field of interest. For an array with maximum baseline B we then find due to Earth rotation an average extrinsic non-planarity $H \sim B/2$ for the longest baseline depending on hour angle range of the observation, declination of the source and latitude of the array. Averaging of the phase deviation

over all baselines leads to a reduced amplitude of less than 1% for point sources at the edge of the FoV as defined above, which is considered acceptable [Taylor, 1999]. More seriously is that the point spread function of these sources will deviate as well. For a compact array with $B = 3$ km we find for $\lambda = 0.2$ m a FoV diameter of 0.007 rad, comparable to the diameter of the beam of a 25m station. However, at a wavelength of 6 m the station beam increases a factor 30 in diameter, while the diameter of the FoV of a 2-D Fourier image increases only by a factor $30^{1/2}$ if we tolerate the same phase error.

This has led to the so-called polyhedron or facet imaging procedure requiring reprocessing of the visibility data by a factor 30 to cover the full FoV with a set of 30 small 2-D Fourier images [see Perley in Taylor, 1999]. For longer baselines than 3 km the number of facets increases linearly and Dutch LOFAR with baselines effectively up to 90 km would even need 951 facets when the 32 m LBA station configuration is used at 50 MHz (table 3.3), which does not seem practical.

It is important to realize that this facet imaging procedure is also a way to implement the corrections to visibilities for each facet for the ionosphere induced position dependent shift and blur of sources, while avoiding imaging errors is a bonus that otherwise should be handled separately [Cornwell, 2008].

2.3.4 Intrinsic array planarity versus extrinsic baseline planarity

If we just consider a reference plane defined by the array itself, we only deal with the Earth curvature where $H = L^2 / 2R_E$ with Earth radius $R_E \sim 6371$ km, which leads for stations at a distance $L = 45$ km from the centre of the array to a non-planarity $H \sim 160$ m. This *intrinsic* non-planarity of an almost planar array is much smaller than the *extrinsic* non-planarity of baselines that appears when baselines between stations on a rotating Earth are observed from outer space over a longer period.

For a reference plane perpendicular to the source direction the non-planarity is defined as the projection of the baseline on this direction. This so called extrinsic non-planarity varies when a sky source is tracked and has for a circular array a largest value defined by longest baseline and source elevation. In contrast, the situation with intrinsic non-planarity has the largest phase deviations on baselines between stations at the centre of the array and stations furthest out.

For a snapshot image with the Dutch LOFAR array and assuming that (2.2) is still valid if the field centre of a small Fourier image is phase shifted to a direction of interest we find at 50 MHz a FoV with diameter of 7° . This would require a station diameter of ~ 59 m if faceting is to be avoided.

Instead of 951 small facet images it seems potentially attractive to form a much smaller set of large snapshot images that can be simply corrected for the varying

shape of the beam and its polarization characteristics. Adding these snapshot images together requires appropriate corrections for rotation and re-scaling of each differently projected sky images, which is also a straightforward computational process [Wijnholds, 2005]. This snapshot approach could therefore in principle lead to processing efficient wide-field synthesis image forming at LOFAR frequencies and will be further considered in chapter 3.

2.3.5 Polarization correction in the image

An important aspect of correcting a whole image for polarization effects is that every source in a snapshot image is properly corrected for its local polarization effect, but also the side lobes of sources centred at different locations. Therefore, the side lobes do not get the same polarization correction as the source from which they emanate. When all strong sources are completely subtracted from the observed visibility data including their observed polarization, we assume that the polarized side lobes of all remaining sources average out to the mean beam polarization structure over the field.

Fortunately, the polarization change over a station beam is small as will be shown in chapter 3 and in that case the average side lobe polarization will be almost the same as the beam polarization of the sky noise. Polarization of the receiver noise is caused by cross-talk between receiver chains that is however less than -60 dB and will be too weak to be observed.

2.3.6 Deconvolution problem for synthesis imaging with a changing station beam

When snapshot images are to be combined into a single synthesis image we need to deal with two effects, (i) each snapshot grid has a different grid in sky coordinates and (ii) each snapshot has a different station beam pattern.

Correcting each snapshot image for its instantaneous beam shape and rescaling its source coordinates to proper sky coordinates before co-adding in a synthesis image has as a consequence that the pattern of the array point spread function (psf) around each source becomes different for each source in the field. The synthesis image that is the weighted sum of all corrected snapshots then also gets for each source a different array psf pattern in the sky coordinates.

The beam pattern varies strongly over the field and causes the largest deformation in the psf as function of source position. This effect could be avoided if all snapshots are combined without correction for the amplitude shape of the average beam pattern. However, coordinate rescaling is still needed and also the relative polarization distribution that is different for each snapshot needs to be corrected to avoid depolarization. After these corrections we are in the same situation as with conven-

tional synthesis imaging and the average array psf will be different at each position in the station beam since the different station beam shape of each snapshot defines a different set of weights for every source. This means in the first place that deconvolution by subtraction of a single psf for the whole image field will only be effective for the nearest side lobes, which are affected less by the varying position scaling.

One of the consequences is that more sources have to be subtracted from the visibility data to eliminate the source and its psf artefacts from an image to reach the thermal noise in an image as will be discussed in chapter 5.

Alternatively, the station beam could be controlled to maintain a more fixed shape [Hamaker, private communication] during a synthesis observation, which is in principle possible with a phased array station, and might be an option for source fields that would suffer too much from all residual effects after application of first order corrections.

2.4 Processing issues for imaging, correlation and beamforming

Observing with 64 LOFAR stations provides a set of ~2000 interferometer baselines each with at most ~100,000 frequency channels of ~1 kHz width for 4 polarizations in each correlation integration time. The longest baselines of ~1,200 km need these narrow band channels as well as an integration time of ~0.1 s to avoid signal degradation for sources at the edge of the field as determined by the wide station beam. Combining these figures leads to a potential correlation output rate of complex visibility samples that is ~8 gigasamples per second (Gsample/s).

2.4.1 Data output rate of correlation processing is a bottleneck for European LOFAR

The aggregate correlation input rate from 64 HBA stations with 2 polarization channels of 96 MHz bandwidth is ~12 Gsample/s for complex sampling. If we consider that these input samples are 2 x 8 bit while the output samples are 2 x 32 bit, we conclude that a correlation system for LOFAR could even expand the data rate from ~192 Gbit/s at the input, to ~512 Gbit/s at the output when sampling at 10 Hz is needed. This is in contrast with other synthesis telescopes where the data rate is reduced thanks to larger channel bandwidth and longer integration times that are allowed for the smaller FoV at higher frequencies.

2.4.2 Correlation processing power as reference for processing platforms

So-called FX correlation uses Fourier transformation (F) of the input signals to provide narrow band channels that are cross-multiplied (X) and integrated. For the large number of stations used for the LOFAR FX correlation we can practically ignore Fourier processing in the total computational load. In the sketched situation for every integrated output sample with bandwidth δf and integration time δt we need a correlation processing capacity $\delta f \delta t \sim 100$ Complex Multiply Add (CMA) operations, where each CMA takes 6 floating point operations (flop). We now have a reference for the number of CMA operations required by the image forming to get an impression of the size ratio between correlation processing platform and image forming platform if imaging has to keep up with correlation.

The total processing power in flop/s required for correlation follows from the total number of baselines (~ 2000), number of polarizations (4), number of channels ($\sim 100,000$) flop per CMA (6) and bandwidth per channel (~ 1 kHz) and is then ~ 4.8 Tflop/s. This number could even be doubled when 2×4 bits are used for the station signals to transport 2 station beams at full bandwidth to the correlation platform [Nieuwpoort, 2009]. Even in that case only half of the available processing power on the correlation platform would be used and additional processing could be contemplated.

2.4.3 Processing for source subtraction and U,V-gridding dominates correlation processing

Source subtraction and gridding of the U,V-samples to a rectangular grid are the most CPU intensive applications in this image forming. The efficiency of this type of processing has been measured for source subtraction using a typical LOFAR dataset and it was confirmed that the overhead in U,V-coordinate evaluation divided by the total number of baseline channels that need to be corrected is small compared to the nominal 3 CMA per source per complex visibility for phase only correction.

The amplitude corrections for time and bandwidth decorrelation effects can be estimated to first order to contribute another 3 CMA, which means that subtraction of typically 400 sources requires at least 2400 CMA per visibility sample. This should be compared with complex gridding with a typical 10×10 kernel that requires only 100 CMA and both operations together would require 2,500 CMA per visibility sample. Since at least two passes are needed in the conventional iterative image processing approach we need at the image forming platform at least 5,000 CMA per output sample from the correlation platform that needed only 100 CMA for that sample.

The conclusion is that full FoV imaging of a single station beam at the highest spatial resolution of LOFAR would require a general purpose processing platform with 50 times the processing power used by the correlation platform if imaging has to keep up with correlation. This dramatic conclusion puts focus on the large number of subtractions that is assumed and will be further addressed in chapter 5.

2.4.4 Full Field-of-View can be handled in principle with dedicated imaging procedures

The Dutch LOFAR array with baselines up to 90 km needs only 10 kHz channels and 1 s integration time, which reduces the input data rate as well as the processing power of the imaging platform by a factor 100. The processing capacity for image forming will even then already require half as much processing power as the cross correlation just to keep up in real time and is dominated by the source subtraction process if indeed an average of some 400 sources need to be subtracted. If this processing could be organized in real time streaming mode, it could in principle even be realized on the existing correlation platform.

For high-resolution imaging with the full European array, we could in principle reorganize the visibility dataset into 100 subsets of 10 kHz channels and 1 s integration time, each for a facet within the station beam. The total visibility sample rate would stay the same and still be excessive, but it could then be argued that only 4 sources need to be subtracted in each sub field reducing the processing load to 100 CMA for convolution and 24 CMA for subtraction per visibility sample. Since the station beam then no longer works as a Nyquist filter that limits aliasing effects of the FFT imaging, we need to adapt the gridding convolution to make it an effective spatial filter for each facet. The processing load per facet dataset is then reduced by a factor $2500/124 \sim 20$, but the total load for 100 facets is still 5 times larger than the load for imaging with Dutch LOFAR. When we assume that the nominal processing for image forming with Dutch LOFAR power is indeed available then only $1/5^{\text{th}}$ of the total FoV provided by the European array could be processed in real time if 4 subtractions per facet is indeed sufficient.

These examples make clear that new processing schemes for image forming are mandatory to handle the huge visibility output rate of the LOFAR correlation processing on an affordable processing platform. Such new processing schemes will be discussed in chapter 3.

2.4.5 Correlation on a general-purpose platform

Traditionally in radio astronomy, correlation platforms are custom made using dedicated chips that handle data streams from 2-bit digitizers with input bandwidth up to about 2 GHz. LOFAR uses only a maximum bandwidth of 100 MHz, but needs

signal sampling at 12-bit to handle the man-made transmissions in the observing band. After spectral filtering, complex samples are obtained of 2×16 bit that can be reduced to 2×8 bit and even 2×4 bit for the spectral channels that will be cross-correlated and which contain mainly celestial noise. Developing custom chips for this bit range using available technology would not be cost-effective compared to standard chips using 18-bit or even floating point arithmetic but realized in state-of-the-art chip technology. Building complete systems from such commercially available chips is a well-established activity supplying state-of-the-art platforms on a commercial and competitive market. Buying a platform is then cost effective and requires only appropriate programming skills to implement a correlation system. This possibility was further investigated [Schaaf, 2003] whereby a cluster of PCs was identified as a potential High Performance Computing (HPC) platform for cross-correlation. In such an approach a multi-dimensional torus network could be realized and processing capacity enhanced with additional modules such as Graphic Processing Units (GPU). An important realization is that a properly configured commercial routing network [Bregman, 2002] could well do the transposing operation needed in cross-correlation of a large set of narrow signal bands from a large set of antennas.

In the end, it has been decided by the LOFAR project to use a commercial super-computer of unconventional design with the appropriate mix of processing power, memory, external I/O capacity, and internal routing capacity to perform the correlation as well as necessary subsequent corrections before integration into datasets that will be used for further calibration and image forming. In this way an external company could separate the development of correlation software from the development of the High Performance Computing (HPC) platform. Also the HPC platform could in principle be bought as late as possible to maximize the performance for a given budget based on technology development in other parts of our society.

This approach has been used to develop the correlation software on a first generation BlueGene/L system [Romein, 2006] and was then reinstalled on a more power efficient second generation BlueGene/P platform using the latest technology. For LOFAR, a two-step process has been implemented where the 10 GbE trunk lines carrying the 195 kHz subbands from the fields are routed to the BlueGene/P HPC where each 10 GbE input line can handle 7 Gbit/s effectively [Romein, 2010]. The internal torus network of the HPC system does the further routing to nodes for Fourier transformation, and brings the results of each sub-band of all stations together at the appropriate processing nodes for cross-correlation.

The originally to be processed bandwidth per receiver of only 30 MHz has been increased to 48 MHz and in a next step four station beams will be processed simultaneously using 2×4 bit station samples. In this final stage about a quarter of the available input bandwidth and half of the available processing capacity will be used, still leaving enough capacity for additional processing.

2.4.6 Dedicated station processing platforms versus general purpose correlation platform

Station beam forming needs, for each output sample, 24 up to 96 input samples that are added together using a complex weight, while cross-correlation requires each station output sample to be multiplied by a sample from 44 to 70 other stations. The total output data rate of all station beam-formers equals the input data rate of the correlation platform and both platforms require about equal amounts of processing power expressed in CMA/s, which suggests comparable types of signal processing platforms. However, each station beam-former needs about as many input signals as the correlation platform and all these antenna signals need, before beam-forming, to be transformed from a high-speed time series to a stream of sub bands. This beam-forming approach requires a poly-phase filter bank that uses order 10 CMA per receiver sample in addition to 1 CMA for just beam forming.

A poly-phase filter bank for every receptor has the additional benefit that only part of the total available 100 MHz instantaneous bandwidth can be selected for further processing, and available processing power for beam-forming can then be used to form additional station beams that increase the instantaneous FoV. Even the full sky could then be covered with limited bandwidth but with selectable frequency, which is highly attractive for a number of non-imaging applications.

Alternatively, a true time delay beamformer could be contemplated, which would require different digital technology that at the time of the design had comparable cost.

In view of the different internal signal routing schemes required for beam-forming and for cross-correlation it has been decided [Schaaf, 2003] to implement a first level of spectral filtering at receptor level and combine it with beam forming on a dedicated station processing platform [Gunst, 2005]. The second level of spectral filtering is implemented at an off-the-shelf High Performance Computing (HPC) platform that uses GbE input from the stations and dedicated internal routing facilities that provide highly efficient cross-correlation processing [Romein, 2006].

2.5 New Considerations in the Design of LOFAR

The ionosphere has a low frequency transmission cut-off around 10 MHz and induces wave front disturbances that define the basic limitations for aperture synthesis imaging at low frequencies. Dealing with those disturbances requires not only innovative calibration [Noordam, 2000], [Noordam, 2006], [Nijboer, 2006], [Tol, 2007], [Yatawatta, 2009], [Tol, 2009], [Smirnov, 2011], but also imaging algorithms that are still under development [Intema, 2009b], [Kazemi, 2011]. For most we need sufficiently large stations with beams that are not too much wider than the scale size of the wave front disturbances and have sufficient sensitivity to observe a

number of self-calibration sources per beam per ionosphere coherence time [Wijnholds, 2011]. The complex gain factors for at least 5 directions per station beam can then be used to find interpolated corrections for all other sources in the field, an approach that will be discussed in chapter 4.

The following subsections describe the new paradigms that make a synthesis array of large phased array antenna stations affordable.

2.5.1 Short dipole

When a dipole antenna has to be used in transmit mode effective power transfer is required. Since the impedance of the dipole is only real in a small frequency band around resonance when the length is about half a wavelength, impedance matching is then simple and results in a typical relative bandwidth of 10%. At longer wavelength, the dipole is relatively short and its impedance gets a dominating imaginary component. Effective power transfer then requires a matching network that further reduces the effective bandwidth. However, common experience is that a short wave radio receiver works perfectly well with a piece of wire much shorter than half a wavelength.

Electromagnetic theory indeed shows that a simple dipole antenna above a ground plane, when operating below resonance, has a beam pattern that is almost independent of frequency and has an effective collecting area that is proportional to wavelength squared, which is much greater than its physical size. If the effective antenna noise temperature can be made lower than the sky brightness temperature [Ardenne, 1999], which increases only slightly steeper with wavelength, then the sensitivity of such an antenna is almost independent of wavelength for most sky sources [Bregman, 1999]. Although proper power matching to a low noise receiver is not possible over a wide frequency band, the effective receiver noise contribution can be made lower than the sky noise over more than an octave bandwidth for frequencies below 240 MHz [Tan, 2000].

This realization turned the short dipole, known by electronic engineers as a narrow band transmit- device, into a wide band receive element, and found implementation not only in LOFAR but also in other astronomical low frequency applications. It meant a paradigm shift leaving the huge fat dipoles of the UTR-2 (Ukraine) as an artefact of an era when low noise transistors were not available.

2.5.2 Station configuration with expanding shells

A regular array with a fixed spacing between the element antennas equal to half the longest wavelength will suffer from grating lobes at shorter wavelengths. A fractal distribution of elements with shorter spacing between elements that are closer to

the centre of the station allows a frequency dependent taper that reduces at higher frequencies the contribution of elements that are further out. These outer elements would reduce the main beam width and give high side lobes since they have larger separation to support their potential collecting area at the lower frequencies. In this way grating lobes are avoided and the station beam can also be made independent of frequency while the aperture efficiency of the station could be about 50% over more than two octaves frequency range [Bregman, 2000b]. This approach has finally led to a configuration with linear expanding annuli comparable to the exposhell with exponential expanding annuli used for the array configuration [Bregman, 2005]. Each annulus has an equal number of elements that are uniformly distributed also in relation to adjacent annuli. By selecting the appropriate half of the available elements, such a configuration has limited effective area sparseness over a specific semi-octave of frequency range [Cappellen, 2004]. In this way the most expensive part of the station, the digital receiver, is used more effectively and allows for each semi-octave to be observed with an effective aperture efficiency of order 50%.

2.5.3 Calibratability, image forming & processing

The calibration and imaging limitations encountered with conventional processing approaches have already been mentioned in subsections 2.1.3 and 2.3.1. It was shown in subsection 2.3.2 how these could be addressed by designing an array with a sufficient number of stations that are large enough and have sufficient bandwidth to allow adequate self-calibration of ionosphere induced wave front distortions. In subsection 2.4.4 we showed how appropriate image forming software would require processing power comparable to that needed for correlation. These aspects are generic for any synthesis array observing at low frequencies, but an array that uses stations that are also phased arrays needs to handle four additional issues:

- grating lobes,
- the so-called blind scan angles,
- changing beam shape by foreshortening and
- changing polarization properties.

These issues are considered in following subsections.

2.5.4 Grating lobes & blind angles

Phased array stations with a regular array of tiles that have a regular but sparse antenna grid show not only grating lobes but also so-called blind angles. Grating lobes produce additional station beams that appear above the horizon if the main beam is pointed below a certain elevation that depends on the observing frequency.

A blind scan angle means that the received signal at a specific frequency is strongly reduced for a specific direction that is determined by the mutual coupling impedances of the antenna elements in an array and by the input impedances of the receivers. Quite fortunately, for sky noise limited receiver systems the reduction in received source power is to first order compensated by a reduction in sky noise power [Cappellen, 2006]. The blind angles have a scale size that is not much larger than that of a station beam and could cause 50% signal drop when the tracking station beam passes through the relevant angle; moreover they have only limited bandwidth.

The directions of grating lobes and blind scan angles are coupled to the configuration of the array, so their effect on a synthesis image can be reduced by orienting the stations differently in the plane of the synthesis array [Wijnholds, 2008], [Bregman, 2012]. To eliminate the effect of a specific station grating lobe when it just passes over a strong sky source or of a blind angle when the main beam passes through, we can just delete all interferometers that share such a station.

In LOFAR, we do take care that each station has a different orientation, so there is no sky source that will be missed in any snapshot observation, so deleting the visibilities associated with a single telescope has only minor sensitivity loss as a consequence.

2.5.5 FoV pattern of a snapshot image defined by the average over all station beams

The effective beam pattern over the FoV of a snapshot image is some average over the beams of the stations. In fact we have an average over all baselines that get all a different weight in the image. This view explains why small differences between station beams lead to distortion of sources depending on their location in the beam since each source get a different weight over its baselines.

The beam for a specific interferometer is the product of the voltage beam patterns of the two stations that form the interferometer. When the configuration of a phased array station is rotated in the plane of the synthesis array but differently for each station then the average side lobe pattern will be reduced, especially the grating lobes and blind angles [Bregman, 2005]. Consequently, all responses of sources outside the main beam will be reduced strongly. By averaging all the snapshots in a synthesis observation, these side lobes rotate over the sky and their spurious responses are reduced even further. It should however be realized that in every instantaneous snapshot certain baselines could observe an object that is positioned in a strong grating lobe of one telescope and in a much weaker side lobe of the other telescope. If the source is sufficiently strong appropriate self-calibrated parameters could be obtained and the source could be properly subtracted. It turns out that only a few sources in the sky are strong enough to reach this subtraction

level and that the remaining ones are much weaker such that their residual effects can be ignored [Wijnholds, 2008].

2.5.6 Snapshot corrections for beam shape and polarization

The snapshot images need correction for beam polarization since it has to be realized that it is the average beam of all dual polarized receptors in a phased array station that defines a fixed large-scale polarizing pattern that moves over the sky by Earth rotation. When a station tracks a field in the sky, its main beam selects a specific part of this large-scale polarizing pattern of the element beam that has only a few percent variations over the piece of sky selected by that main beam.

To simplify the forming of a polarized image with a synthesis array, it has been decided to orient all antenna elements of all phased array stations identically in the plane of the synthesis array as far as possible [Bregman, 2012]. This means that each snapshot image needs the same polarization corrections for all of its baselines before co-adding into a synthesized sky image and allows correction to be applied before as well as after Fourier transformation of each snapshot (see chapter 3 for more detail).

However, differential Faraday rotation by different ionosphere thickness over the stations and by different Earth magnetic field strength and direction, needs to be corrected per baselines and will be further discussed in chapter 4.

2.5.7 Expo-shell array configuration

The spatial distribution of the stations defines the brightness sensitivity at all resolution scales of the instrument. For an array like LOFAR where we cannot change the configuration to match an observational brightness sensitivity criterion, as can be done for instance with the movable antennas of the VLA, we need a configuration where appropriate subsets of visibilities could be selected after the observation. This results in images with reduced collecting area and point source sensitivity but with a brightness sensitivity that is properly matched to the observational requirements.

For LOFAR, a so-called exponentially expanding shell concept has been adopted [Bregman, 2012] where each annulus has an equal number of ~ 5 stations that are uniformly distributed, also in relation to adjacent annuli. This leads to a U,V-distribution with shells that contain 10 points that could be extended in radial direction if sufficient bandwidth is used. Earth rotation could then give full coverage within ~ 3 hours for sources with appropriate declination.

2.5.8 Summary of paradigm shifts

The most important paradigm shifts are summarized as follows:

- 1) Large sparse phased arrays stations have an effective collecting area that varies with frequency and gives first order compensation for the frequency dependent sky noise that dominates the system temperature [Bregman, 2000b].
- 2) Electrically small dipole antennas make large stations affordable [Tan, 2000].
- 3) Design strategy relying on Moore's law allowed a start of design in 2001 based on the expected performance of signal processing components that are ordered after 2003 [Bregman, 2000a].
- 4) Array configuration based on Exponential shell distribution [Bregman, 2005].
- 5) Non-identical station configurations optimize array performance by rotation of the configuration while the orientation of the antenna elements is equal [Wijnholds, 2008], [Bregman, 2012].
- 6) Off-the shelf platform for cross-correlation allowed porting of the correlation software to a next generation platform within a couple of months [Romein, 2010].
- 7) Calibration procedures that model the ionosphere phase screen and extract parameters that allow accurate subtraction of the strongest set of sources that would otherwise determine the effective noise in an image [Noordam, 2000, 2006], [Nijboer, 2006], [Tol, 2007], [Yatawatta, 2008], [Intema, 2009b], [Smirnov, 2011], [Kazemi, 2011], [this dissertation].
- 8) Imaging procedures that handle polarization over the large field-of-view provided by the element antennas in the stations [Hamaker, 2000], [Yatawatta, 2012a], [this dissertation].

It is important to realize that all paradigm shifts that allowed realization of LOFAR have all been made possible by 21st century technologies that allow efficient digital processing of the signals of large sets of wave front sensors. Even more important has been the development program initiated at ASTRON that brought these potential technologies to the field of radio astronomy [Ardenne, 1997], [Ardenne, 1999], [Ardenne, 2000], [Ardenne, 2002].

3 **Efficient Processing for Wide-field Synthesis imaging**

At the conception of LOFAR [Bregman, 1998; 1999], it was realized that existing self-calibrating imaging packages could not cope with the large field-of-view (FoV) that would be provided by antenna stations that operate at low frequencies. This was especially true for arrays with phased array antenna stations, where the beam changes shape during an observation, requiring new imaging procedures for which a large number of important aspects will be discussed in this chapter.

This chapter reports a study on the limitations to the FoV of Fourier imaging with a non-planar correlation array, and more importantly, it shows all the basis ingredients of efficient processing for wide-field imaging.

Global introduction

Conventional Earth rotation synthesis with 2-D arrays uses mostly imaging with a single Fourier plane, which involves projection to handle the observed baseline visibilities that span a volume due to Earth rotation. This projection leads to a field-of-view (FoV) much smaller than the station beams of LOFAR and is defined by phase errors on the longest baselines for objects at the edge of the FoV. There are various ways to reduce these phase errors to low levels such that object distortions and additional image noise are acceptable.

Most imaging packages for 2-D aperture synthesis arrays handle the inherently limited accuracy of approximate 2-D Fourier imaging by using an iterative imaging process often combined with self-calibration [Taylor, 1999]. In this chapter, however, we focus on the disturbing phase terms that arise when a 2-D Fourier transform is used to form an image with a non-coplanar set of interferometer baselines.

Even more importantly, current implementations of the various methods ask for too much processing power to be of practical use for LOFAR, requiring a more efficient processing approach. An important concept is to distinguish between intrinsic non-planarity of baselines when stations follow Earth curvature, and extrinsic non-planarity in the baselines of a long observation with a planar array that are induced by Earth rotation. This separation in origin is important because of the different methods of dealing with these effects. Since the central core of the LOFAR array is almost flat, a single 2-D Fourier transformation could provide in principle a FoV that covers a hemisphere with only minor distortions. Instead of handling a set of baselines that span a volume by Earth rotation, we need to handle rotation of the sky by a set of snapshot images.

This snapshot approach was successfully demonstrated to synthesise a large sky image from data obtained with the LOFAR initial test station covering even more than a hemisphere [Wijnholds, 2004]. The approach is a useful tool to analyse the effects of changing beam shape of a phased array station on a synthesis image. It evolved during analysis of residual imaging errors that remain after phase correction for second order distortions that arise in 2-D Fourier imaging with non-planar arrays. Finally it turned out that the snapshot approach is not only a simple analysis tool but that third order phase errors can be kept sufficiently small to make it potentially a high efficient processing approach for wide field imaging with the Dutch LOFAR array.

An attractive feature of the snapshot approach is that individual images cannot only be simply corrected for foreshortening of the beams of phased array station but also for the polarization of the element antennas in these stations.

Application of this snapshot method to configurations that also include the European stations would however require faceting of individual snapshot images, and raised the question whether the conventional faceting approach could be made more processing efficient. We indeed found such a solution coined Fast Faceting, where just as in a Fast Fourier Transform (FFT) all possible facets are made without increasing the total data volume. If implemented at the correlation platform, only a subset of the huge data volume could be chosen for actual imaging to limit the output data rate and processing requirements for imaging.

Background

The starting point for this analysis is the Van Cittert-Zernike theorem that relates the observed spatial coherence (or visibility) function to the brightness distribution of the incoming radiation. The theorem shows that a Fourier Transform of the brightness distribution can describe the spatial correlation function if certain conditions are met. Proof of the Van Cittert-Zernike theorem can be found in a number of textbooks such as “Principles of Optics” [Born, 1999] and “Interferometry and Synthesis in radio astronomy” [Thompson, 2004]. An important aspect of Fourier imaging is that it gives a simple description only for a single frequency, which is approximately valid for a small relative bandwidth. When a larger bandwidth has to be handled, as in continuum imaging, we need to combine a set of images where the side lobe pattern around each object scales with frequency and where different objects vary differently in intensity as function of frequency, effects to be dealt with in so called multi-frequency synthesis [Rao, 2010]. We continue the introduction of this chapter with a summary of the Fourier based imaging approaches currently in use for Earth rotation synthesis and we conclude this introduction with an outline of the chapter.

We assume that all sources are at great distances from the interferometer such that plane waves are received from each direction. Direction vectors can be projected

on a plane, and if all observed visibilities lie also in this plane a simple 2-D Inverse Fourier transform describes a full sky image that is the sum of two hemispheric projections. A reference Cartesian system plane can be chosen to best fit the actual physical circumstances of an array. E.g. for an Earth rotation synthesis array with only East-West baselines, a coordinate system with a reference plane perpendicular to the Earth polar axis will be best, since all rotated baselines lie in that plane.

The non-astronomical community uses the plane of a 2-D array as the reference plane for making 2-D snapshot images that contain only information from one hemisphere, since the Earth shields the other hemisphere.

Due to Earth rotation, snapshots of a part of the sky have a changing orientation and different foreshortening depending on the elevation of the FoV with respect to the array plane. Corrections for these effects have to be applied before individual snapshots can be added. After these corrections the point spread function (psf) will vary over the observed FoV, making deconvolution procedures that assume a constant psf impossible. As a result, past implementation effort for long synthesis observations has been concentrated on alternative imaging approaches such as 3-D Fourier inversion and polyhedron imaging, as discussed by Perley [chapter 19, Taylor, 1999] starting from a 3-D Cartesian reference system.

The 3-D Fourier imaging approach transforms the 3-D visibility data cube in U,V,W-space into a data cube of intensities in l,m,n-space and finds the image on the unit sphere of direction cosines defined by the constraint $l^2 + m^2 + n^2 = 1$. When the n-axis is chosen towards the centre of the source field, only a small volume needs to be transformed that contains the surface of a spherical cap. Although conceptually simple in explaining synthesised sky imaging with a set of Earth bound interferometers that rotate relative to the sky, it gives no simple answers to important questions such as how non-stationary sources appear in the final image or on the effect of a varying foreshortened beam of phased array stations. More serious is that 3-D imaging requires a set of 2-D Fourier planes where each plane covers the full extent of the FoV but all planes together need to fill the whole volume of the spherical cap. Compared with a single plane, a processing penalty is involved that is proportional to the FoV of the spherical cap and to the longest baseline in wavelengths. LOFAR has a large FoV when stations of 32 m diameter are used at 50 MHz, and would need ~400 planes for imaging with baselines up to 120 km, which requires more Fourier processing power than can be afforded.

Polyhedron imaging is an extension of the conventional 2-D Fourier approach where the baseline volume is projected on the plane of the image. Phase deviations are limited by reducing the imaged field extent by a convolution of the visibility data. To image the large FoV of the main beam of the array stations a number of smaller distortion free 2-D Fourier images are required. Application to LOFAR would require a large number of small facets. Unfortunately, current implementations reprocess all

visibility samples for each facet image, which leads for LOFAR to more processing power just for data inversion than can be afforded.

A recent method called W-projection [Cornwell, 2008] corrects for the phase errors in 2-D Fourier imaging due to W-terms in a synthesis observation and tries to avoid partitioning in facets by applying a complex quasi-convolution to the measured visibility data prior to Fourier transformation. Unfortunately, as will be shown, the linear extent of the required convolution kernel scales proportionally to the FoV and to extrinsic non-planarity, which leads for LOFAR to more pre-processing power for the visibility data than can be afforded.

Approach

We start our analysis from first principles and arrive at a number of results that have great practical consequences. The most important one is a detailed analysis of the fringe shift theorem for Fourier transforms in 3-D and 2-D that is based on invariance of the vector product for rotation of the coordinate system. A planar configuration in 3-D then reduces simply to 2-D by rotation of the coordinate system. However, complication arises when an intrinsic 3-D configuration is projected to a 2-D one. The conventional approximation will be extended and forms the basis of the proposed synthesized snapshot imaging approach as an alternative for existing synthesis imaging methods.

The first practical contribution is the derivation of the size of a complex convolution kernel that enhances the field of view (FoV) of 2-D Fourier inversion of a non-planar correlation array.

The second contribution is the design of a new method here coined Fast Faceting that allows efficient generation of a large number of small datasets although only a small fraction of the facets need actually to be imaged. This method can be combined with existing synthesis methods and allows efficient processing of very high resolution images as will result from the large extent of LOFAR.

The third contribution, the synthesized snapshot approach, is a new combination of well-known principles and is particularly useful to include aspects of phased array stations, such as foreshortening in the station beam of phased array stations as used in LOFAR and as planned for the SKA. The method also simplifies analysis of long synthesis imaging by describing it as a sum of simple 2-D Fourier images, where each image has its own imaging and calibration artefacts. These artefacts can be described by amplitude and phase deviations in the visibilities that cause side lobe structure around sources depending on their location in each image.

The chapter is partitioned to a number of sections with subsections and detailed conclusions.

Section 3.1 gives an outline of the Van Cittert-Zernike theorem forming the basis for our analysis. We show how actual phased arrays that have deviations from planarity give distorted objects in a 2-D Fourier image, which limits the useful FoV of that image.

Section 3.2 describes how integration in time and frequency by a correlation interferometer that tracks moving sources causes degradation effects that limit the FoV around such sources, and have a serious impact on the required processing capacity for wide-field image forming at high resolution.

Section 3.3 describes the effects of the convolutional re-gridding of observed interferometer data to a rectangular grid such that the processing efficient Fast Fourier Transform (FFT) can be used for imaging. Special attention is paid to the spatial filtering by such a convolution to limit the FoV such that aliasing artefacts of the FFT are reduced to acceptable levels.

In section 3.4 we analyse current approaches to extend the FoV of a 2-D FFT image, such as complex convolution correction of the station based non-planarity effects, as well as separation of the total FoV into a number of smaller facets.

Section 3.5 analyses the 2-D Fourier snapshot imaging approach that uses an array based coordinate system for a quasi-planar array with only limited intrinsic non-planarity. A procedure is derived that maximizes the tracking time when a rotating sky field is tracked.

Section 3.6 describes the effects of polarization in a station beam as induced by the element beam of the antennas in a phased array station. Also, the so called blind angles in the average element beam and their impact on the station beam are discussed as well as mitigation strategies to reduce these effects together with the effects of station grating lobes.

Section 3.7 compares the processing aspects of various imaging approaches. We show in the first place that for continuum imaging, such as for LOFAR, the processing is dominated by the source subtraction process if more than about 20 sources have to be subtracted.

Section 3.8 discusses how signals from other directions appear in a synthesis image where U,V-coordinates are rotated and visibilities are fringe shifted to correct for sky rotation when a sky field is tracked.

Section 3.9 summarizes the conclusions of individual sections in a broader perspective.

3.1 Field-of-View of 2-D Fourier imaging with a non-planar array

We start this section with consideration of the Van Cittert-Zernike theorem as implemented with the so-called Measurement Equation that describes the response of a single interferometer. We show how a 3-D distribution of baselines can be handled by a 3-D Fourier inversion but that a planar array only needs 2-D inversion. There is a shift theorem for 2-D and for 3-D Fourier transforms, but the 3-D shift between two projected 3-D images is not position invariant. Non-planar phased arrays show phase errors in their visibilities that are baseline and source position dependent when 2-D Fourier inversion is attempted. These phase errors distort an observed point source as function of its position after 2-D Fourier inversion. As a result, also the side lobe pattern appearing to emanate from this source becomes distorted. Although the nominal side lobe pattern can in principle be removed by subtracting the response of a nominal point source from the image, the residuals of the distorted side lobes introduce a noise background proportional to the strength of the source.

We introduce a limit for the FoV of the snapshot image as determined by the effective reduction in amplitude of objects at some distance from the centre of an image due to phase deviation in the visibilities. An important insight is the distinction between intrinsic non-planarity of the array caused by Earth curvature and extrinsic non-planarity as created in most current legacy image forming packages using 2-D FFT processing. Finally we show how a combination of model fitting and direct inversion can lead to images where residual artefacts are reduced to acceptable levels.

3.1.1 Basic Interferometer Measurement Equation

We start our analysis with the cross-correlated response of two antenna (stations) that form an interferometer. We assume that all sources are at great distance from the interferometer such that plane waves are received from each direction.

An antenna receptor with its phase reference centre defined by position vector \mathbf{r} in a Cartesian x, y, z -coordinate system has an associated spatial frequency vector \mathbf{u} with coordinates $u = x / \lambda$, $v = y / \lambda$ and $w = z / \lambda$ where λ is the wavelength of the signal with frequency ν given by $\nu = c / \lambda$ where c is the speed of light.

For an object in a direction defined by unit vector \mathbf{l} with direction cosines l , m and n relative to x -, y - and z -axes respectively, the geometric antenna phase ϕ is defined relative to the origin of the coordinate system by the following vector product

$$\phi / 2\pi = \mathbf{u} \cdot \mathbf{l} = (u l + v m + w n) \quad [\text{rad}] \quad (3.1)$$

The complex voltage response V_{ik} of an antenna i with effective electrical length l_{ei} and normalized voltage beam response g_{ik} on a plane wave from direction \mathbf{l}_k with electric field strength E_k for monochromatic radiation that is polarization matched to the antenna is given by

$$V_{ik} = l_{ei} g_{ik} E_k \exp(-i \varphi) \quad (3.2)$$

where the minus sign in the exponent is a matter of convention and i is the square root of -1.

The correlated response c_{ijk} of two antennas i and j on the plane wave from direction \mathbf{l}_k is given by

$$c_{ijk} = \langle V_{ik} V_{jk}^* \rangle_{\delta t \delta \nu} \quad (3.3)$$

where $*$ indicates complex conjugation and $\langle \rangle_{\delta t \delta \nu}$ indicates averaging over time interval δt and spectral channel bandwidth $\delta \nu$ by the correlation processing of a noise signal characterized by spectral power density. Insertion of (3.1) and (3.2) in (3.3) gives

$$c_{ijk} = \langle l_{ei} g_{ik} E_{ik} \exp(-2\pi i \mathbf{U}_{ij} \cdot \mathbf{l}_k) E_{ik}^* l_{ej}^* g_{jk}^* \rangle_{\delta t \delta \nu} \quad (3.4)$$

where the spatial frequency vector \mathbf{U}_{ij} of the interferometer formed by two antennas at positions \mathbf{r}_i and \mathbf{r}_j is given by $\mathbf{U}_{ij} = \mathbf{u}_i - \mathbf{u}_j$. The common origin of \mathbf{u}_i and \mathbf{u}_j has now dropped from the equation and \mathbf{U}_{ij} is the baseline between the phase reference points of the two antennas, which will be further discussed in subsection 3.6.1.

We assume that averaging of all variables, parameters, and exponentials in (3.4) over narrow band $\delta \nu$ and the small interval δt can be absorbed by a small reduction of the signal amplitude, which will be discussed in section 3.2.

The measured response c_{ij}^m of a correlation interferometer is a summation over a set of K source contributions c_{ijk} and if each source provides an incoherent complex power contribution, we find

$$c_{ij}^m = S_{ij} \sum^K c_{ijk} \quad (3.5)$$

where the sampling function S_{ij} assigns each correlated response value to a point \mathbf{U}_{ij} in visibility space corresponding to the centre of the integration interval and the centre of the spectral channel.

The single voltage equations for a single polarization can easily be generalized to full polarization by replacing the scalar product in (3.4) by a 2x2 Jones matrix product [Hamaker, 1996]. All factors are already placed in proper order and instead of the scalar complex conjugation operation $*$ we need the matrix operation H for com-

plex conjugation and transpose. The matrix equation then describes the 4 polarized coherency components of the sources and observed visibilities and the receptor gains become direction dependent Jones matrices [Smirnov, 2011] for each dual polarized receptor pair.

Inserting (3.4) in (3.5) and using the interferometer form of (3.1) with $\mathbf{U}_{ij} = (U_{ij}, V_{ij}, W_{ij})$ we get

$$c_{ij}^m = S_{ij} \mathbf{l}_{ei} \mathbf{l}_{ej}^* \sum^K g_{ik} g_{jk}^* \langle E_{ik} E_{jk}^* \rangle_{\delta t \delta v} \exp(-2\pi i (U_{ij} l_k + V_{ij} m_k + W_{ij} n_k)) \quad (3.6)$$

The reordering in (3.6) implies that c_{ij}^m and $\langle E_{ik} E_{jk}^* \rangle$ are coherence vectors of length 4 and the products $\mathbf{l}_{ei} \mathbf{l}_{ej}^*$ and $g_{ik} g_{jk}^*$ have become 4*4 Mueller matrices that convert the observed coherencies to Stokes parameters.

For a plane wave we replace the averaged field correlation $\langle E_{ik} E_{jk}^* \rangle_{\delta t \delta v}$ by the received power density $P_k \delta\Omega_k \delta v$ from direction \mathbf{l}_k and we get

$$c_{ij}^m = S_{ij} (\mathbf{l}_{ei} \mathbf{l}_{ej}^*) \sum^K \delta\Omega_k \delta v g_{ik} g_{jk}^* P_k \exp(-2\pi i (U_{ij} l_k + V_{ij} m_k + W_{ij} n_k)) \quad (3.7)$$

The summation over K solid angle elements extends over the full sky, where P_k is the average power density over solid angle $\delta\Omega_k$ and is a measure for the brightness temperature $T_B(\mathbf{l}_k)$ from direction \mathbf{l}_k . In practical systems, we find some normalized value of c_{ij}^m , which requires proper renormalization to find either P_k in proper temperature units or $\delta\Omega_k P_k$ in flux units.

3.1.2 3-D Fourier Inversion

We simplify (3.7) further by defining a single effective antenna aperture $A_e = \mathbf{l}_{ei} \mathbf{l}_{ej}^*$ instead of individual interferometer apertures and absorb differences per baseline by appropriate calibration of the voltage beam products $g_{ik} g_{jk}^*$.

A more important simplification is eliminating a beam dependency on baseline by assuming a single –averaged- power beam response $g_k^p = g_k g_k^* = g_{ik} g_{jk}^*$ independent of indices i and j. Then we get

$$c_{ij}^m = S_{ij} A_e \delta v \sum^K \delta\Omega_k g_k^p P_k \exp(-2\pi i (U_{ij} l_k + V_{ij} m_k + W_{ij} n_k)) \quad (3.8)$$

We now recognize in (3.8) a 3-D Fourier relation between the observed sky brightness $g_k^p P_k$ expressed in $[W m^{-2} Hz^{-1} sr^{-1}]$ and the measured correlated interferome-

ter response c_{ij}^m . However, we need to realize that radiation propagating from a large distance towards our antennas has only electric field components perpendicular to the propagation direction and apparently appears to emanate from a sphere. Consequently we need only image pixels on a unit sphere given by $l^2 + m^2 + n^2 = 1$. This means that if a 3-D Fourier inversion from U,V,W-space to l,m,n-space would be done we need to interpolate the 3-D brightness results on a sphere as discussed by Perley in chapter 19 of [Taylor, 1999]. We finally get $\delta\Omega_k g_k^p P_k$, which is the apparent flux expressed in [$W m^{-2} Hz^{-1}$]

As equation (3.8) is invariant for the orientation of the coordinate system, we could choose the n-axis in the same direction as which where the station beams are pointed. Imaging can now be realized by 3-D Fourier inversion of (3.8), which can be realized as a series of 2-D Fourier transforms instead of a single 2-D transform.

We give a first order evaluation of the number of computations for the 3-D inversion, which is proportional to the number of planes that have to be evaluated. If we need imaging only over a limited circular area $\pi \Delta l^2$ then the image volume of the spherical cap has a height $(1-n) \sim \Delta l^2/2$ so the total volume is $V = \pi/2 \Delta l^4$. An interferometer with maximum baseline B_m gives at wavelength λ a resolution λ/B_m and requires an average pixel separation $\lambda/2B_m$ so the volume of a single pixel is $\Delta V = (\lambda/2B_m)^3$. Imaging of the field covered by the primary beam of a telescope with diameter D requires typically $\Delta l \sim \lambda/D$ and the total number of N_{3D} image pixels for a 3-D Fourier inversion is then given by

$$N_{3D} \sim 4\pi \lambda B_m^3 D^{-4}$$

This is a dramatic result since the final number of image pixels N_{im} on the sphere is only

$$N_{im} \sim 4\pi B_m^2 D^{-2}$$

These two equations tell us that we get at least an excess factor F_{3d} in the required number of computations per image pixel for 3-D Fourier imaging over 2-D Fourier imaging given by

$$F_{3d} = N_{3D} / N_{im} \sim \lambda B_m D^{-2} \quad (3.9)$$

The excess factor is proportional to the total FoV ($\sim \lambda^2 D^{-2}$) and to the inverse resolution ($\sim B_m \lambda^{-1}$) and becomes large for large arrays at long wavelength with small stations.

In practice the situation is less dramatic, since the sampling in the n direction is not determined by B_m but by the range of baseline values in the W-direction. This means that we need to replace in (3.9) B_m by B_p which is the projection of B_m on the

W-axis. A 3-D Fourier transform can be partitioned in a set of 2-D ones and the number of 2-D Fourier planes that fills the 3-D l, m, n -space is therefore given by

$$N_{pl} \sim \lambda B_p D^{-2} \quad (3.10)$$

This formula is consistent with, but differs from (19-16) in [Taylor, 1999] which assumes a larger FoV than used above.

For Dutch LOFAR with $B_m \sim 120$ km, $D \sim 32$ m and $\lambda \sim 6$ m we need $N_{pl} \sim 400$ if we assume $B_p \sim B_m/2$ which itself justifies a search for alternative solutions with a smaller processing penalty.

An attractive feature of the 3-D approach is that it can combine correlations from a U,V,W-space as we get by Earth rotation synthesis providing an exact solution without approximations. At this point it is not clear how a changing beam shape as appears for phased array antenna stations could be corrected. As approach, we could for instance make separate snapshot images and correct each one for its polarized beam shape before co-adding. However, also a simple snapshot image suffers from the same 3-D excess factor, while a simple 2-D transform would be correct for an intrinsic planar array.

3.1.3 Spherical projection

A common approach is describing (3.8) as a projection on an arbitrarily chosen equator plane of a sphere, see Clark chapter 1 in [Taylor, 1999], which leads with $\delta\Omega_k = n_k^{-1} \delta l_k \delta m_k$ to

$$c_{ij}^m = S_{ij} A_e \delta v \sum^K \delta l_k \delta m_k n_k^{-1} g_k^p P_k G_{ijk} \exp(-2\pi i (U_{ij} l_k + V_{ij} m_k)) \quad (3.11)$$

with the so called W-term

$$G_{ijk} = \exp(-2\pi i W_{ij} n_k) \quad (3.12)$$

$$\text{while } n_k = (1 - l_k^2 - m_k^2)^{1/2} \quad (3.13)$$

The summation is still taken over the solid angle of the full sky sphere that may be partially blocked by the ground based antenna elements but is now expressed by l, m -coordinates with constant increments δl_k and δm_k in a plane perpendicular to the n -axis.

In cases where W is small we get $G_{ijk} \sim 1$ and its effect can be ignored allowing a simple 2-D Fourier transform to provide a large image field as will be discussed in subsection 3.1.4. For non-zero W , but for small l_k and m_k as will be discussed in

subsection 3.1.8, we get $n_k \sim 1$ and G_{ijk} can be simply corrected identically for all l_k and m_k by the fringe stopping process as will be discussed in subsections 3.1.6 and 3.1.9. In that case a simple 2-D Fourier transform will also be possible, but provides an accurate image only for a much smaller field.

3.1.4 2-D Fourier inversion of Planar Array responses

In case of a planar array it is attractive to choose the W -axis toward Zenith perpendicular to the x, y -plane of the array instead of the direction of the field of interest. In that case we have $W_{ij} = 0$ so $G_{ijk} = 1$ and we recognize in (3.11) the 2-D Fourier relation between measured correlations and apparent brightness distribution $n_k^{-1} g_k^p P_k$ that can be obtained after the 2-D inverse Fourier transform (IFT) according to

$$(n_k^{-1} g_k^p P_k) * \underline{S}_k = F_n \sum_{ij}^M c_{ij}^m \exp(+2\pi i (U_{ij} l_k + V_{ij} m_k)) \quad (3.14)$$

where M is the total number of visibility samples with indices i and j while F_n is a normalization factor.

A serious limitation of the Fourier inversion is that the resulting apparent brightness distribution is given by the desired distribution $n_k^{-1} g_k^p P_k$, but convolved (indicated by operator $*$) with the point spread function (psf) of the measurement setup \underline{S}_k that is the Fourier inverse of S_{ij} . The result of the side lobe pattern of the psf around each point source in the field is that the stronger sources could mask the weaker ones.

We arrived at a well-known result if the additional assumption holds that all receptor –voltage– beams g_{ik} are indeed equal to g_k , which results in a single power beam g_k^p that limits the FoV of a synthesis observation. Apart from incoherency of the source distribution there are additional constraints such as signals being stationary, but these constraints are fulfilled in most astronomical imaging applications with a stable instrument [Thompson, 2004]. We keep our focus on the requirement of identical beams for (3.8), which is for LOFAR however only approximately fulfilled for reasons that will be discussed at the end of this chapter.

With our particular choice of a reference coordinate system in which the W -term is eliminated for a planar array a simple 2-D Fourier transform provides a full instantaneous hemispheric image without distortions, where all sources including RFI sources appear at their nominal positions. More importantly, we can correct this hemispheric image for the polarized beam shape of the element antennas in our phased array antenna stations. Dealing with a rotating sky needs a series of snapshot images as will be shown in the next subsection and will be further discussed in section 3.5.

The tilted planar array uses a spherical projection as discussed in section 3.1.4 with $n_k'^2 = 1 - l_k'^2 - m_k'^2$ on the same sphere that is given by $n_k^2 + l_k^2 + m_k^2 = 1$ so for the plane $m_k = m_k' = 0$ we find a simple rotation by α_r as depicted in figure 3.1.

This analysis shows that there is no need to do a 3-D Fourier inversion for a tilted planar array, we just need to define a new reference system with U' and V' in the plane of the array and then W' is zero which allows a simple 2-D Fourier inversion. Back projection on the unit sphere adds the n_k' coordinate and allows a simple 3-D vector rotation to provide the image coordinates in any l'', m'', n'' -coordinate system, preferably one connected to the sky. However, we need a large set of 2-D transforms when the U', V' -plane connected to the Earth rotates with respect to the final l'', m'', n'' -coordinate system fixed to the sky.

3.1.6 Phase after a fringe shift correction on correlated signals of a non-planar array

A true planar array with n -axis perpendicular to the array plane has no W -terms. However, an actual array suffers from small station dependent W -contributions for instance due to Earth curvature. The main beam of the antenna stations is pointed towards a reference position l_0, m_0 on the sky and we want to investigate the behaviour of the interferometer phase (3.15) for small distance l_s, m_s of a source at (l_k, m_k) from this reference position by insertion of $l_k = l_0 + l_s, m_k = m_0 + m_s$ and $n_k = n_0 + n_s$. Insertion of these values into (3.15) then gives

$$\varphi_s / 2\pi = (l_0 + l_s) U + (m_0 + m_s) V + (n_0 + n_s) W \quad (3.18)$$

We replaced index k in (3.15) by index s , to stress its relation with the shifted coordinates l_s and m_s and dropped indices i and j for notational convenience.

We want however an expression for (3.18) where n_s is eliminated and after simple phase correction of the data a 2-D Fourier transform can be performed to obtain an image centred on (l_0, m_0) .

Define $n_d = n_k^2 - n_0^2$ and eliminate n_k using $n_k = n_s + n_0$ to give

$$n_d = n_s^2 + 2n_0 n_s \quad (3.19)$$

This equation for n_s can be solved and we take the relevant solution

$$\begin{aligned} n_s &= -n_0 + (n_0^2 + n_d)^{1/2} \\ n_s &= -n_0 \{1 - (1 + n_d / n_0^2)^{1/2}\} \end{aligned} \quad (3.20)$$

For small (n_d / n_0^2), i.e. for $n_s \ll \frac{1}{2} n_0$ (as follows from 3.19) we use a series expansion for the square root term and find

$$n_s = n_0 (n_d / 2n_0^2 - n_d^2 / 8n_0^4 + \dots) \quad \text{for } n_s \ll \frac{1}{2} n_0 \quad (3.20a)$$

For a coordinate system with W-axis towards Zenith, the constraint limits the extent of a 2-D FT to stay well above half way between the field centre and the horizon.

Since both l_0 and l_k are located on the unit sphere we can use (3.13) giving

$$\begin{aligned} n_k^2 &= 1 - l_k^2 - m_k^2 \\ n_0^2 &= 1 - l_0^2 - m_0^2 \end{aligned}$$

Inserting these expressions into the definition equation for n_d we find

$$n_d = (-l_k^2 - m_k^2 + l_0^2 + m_0^2)$$

Using $l_k = l_0 + l_s$ and $m_k = m_0 + m_s$ we get

$$n_d = -(l_s^2 + m_s^2 + 2l_0 l_s + 2m_0 m_s)$$

Inserting this result into (3.20a) and ignoring 3rd and higher order terms in l_s and m_s we get after retaining all the terms that are linear and quadratic in l_s and m_s an expression for n_s that can be inserted into (3.18) giving

$$\begin{aligned} \varphi_s / 2\pi &= l_0 U + m_0 V + n_0 W \\ &+ l_s (U - W l_0 / n_0) + m_s (V - W m_0 / n_0) \\ &- (l_s^2 + m_s^2) W / 2n_0 - (l_0 l_s + m_0 m_s)^2 W / 2n_0^3 \end{aligned} \quad (3.21)$$

The first line in (3.21) describes the canonical 3-D fringe shift term. The fringe stopping process applies a phase correction per station and subtracts this phase shift from every interferometer. For a calibrated set of interferometers the resulting phase in (3.21) is zero for $l_s = m_s = 0$ independent of U , V and W and therefore defines the centre of a shifted Fourier image in l_s, m_s -coordinates. These coordinates are not direction cosines themselves but need to be added to (l_0, m_0) as indicated in figure 3.2 before they can be projected back to the unit sphere. Then we find true source coordinate vector (l_k, m_k, n_k) that can be rotated to any required l'', m'', n'' -coordinate system by a simple matrix multiplication.

For a planar array with $W = 0$ we find as required the canonical shift theorem in the l, m -plane of a 2-D Fourier transform as depicted in figure 3.2.

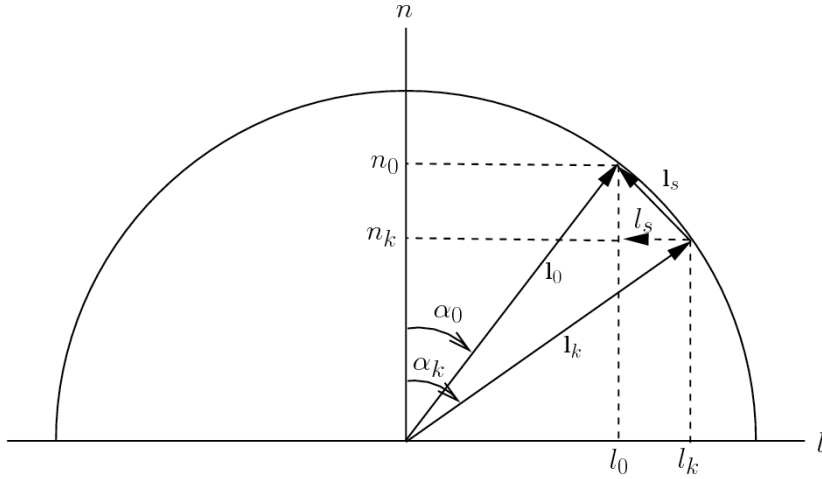


Figure 3.2. Coordinate shifting from l_k by l_s towards l_0 , for $m = 0$ plane.

Equation (3.21) shows that after a fringe stopping correction for (l_0, m_0, n_0) as indicated by the first line, the correlations of a source at $l_k = l_0 + l_s$ and $m_k = m_0 + m_s$ have the correct phase to appear at l_s and m_s in a shifted Fourier image, requiring corrected U', V' -coordinates. These coordinates, according to the second line in (3.21), are given by $U' = U - W l_0/n_0$ and $V' = V - W m_0/n_0$, which are projections of the baseline vector on the reference plane from direction (l_0, m_0, n_0) as indicated in figure 3.3.

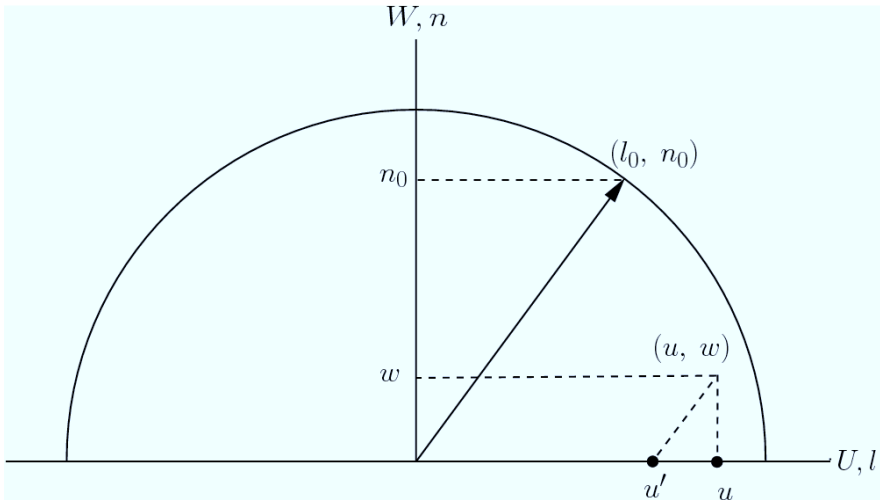


Figure 3.3. W-projection in plane $V = 0$ from direction (l_0, m_0, n_0) on U -axis to get U' as in figure 3.1.

The third line in (3.21) has terms that are quadratic in l_s and m_s but has also a cross term ($l_s m_s l_0 m_0$) that limits the duration of a snapshot image in a coordinate system where l_0 and m_0 change. Corrections for the quadratic terms will be discussed in section 3.4 and limitations by the cross term will be further discussed in section 3.5.

3.1.7 Fringe stopping and fringe tracking

An array that tracks a point that rotates with the sky uses a nominal field position defined by the fringe stopping process that applies a continuous phase correction to every spectral component of every measured correlation at every instant for that nominal direction on the sky. We see in (3.21) that the fringe stopping corrects for the continuous change in angle between the baseline vector and the vector of the reference position, which is independent of the reference coordinate system.

The fringe stopping process is realized in practice by applying a combination of discrete delay steps per station accompanied by appropriate phase corrections. The main reason for fringe tracking is to reduce the output data rate of the correlation process. In addition, according to the shift theorem for Fourier transforms the centre of the image could be placed at an appropriate sky position, preferably the same position that is used by the main beam of the stations that track a sky field. The reduced phase rate of visibilities of objects at limited distance from the fringe tracking position allows choosing integration time and spectral resolution in the correlation process such that the output data rate of the correlation process can be limited without unduly attenuating the visibility amplitude of objects in the station main beam. This aspect will be discussed further in section 3.2.

3.1.8 Field-of-view limitation by non-planarity in 2-D Fourier imaging

When a 2-D Fourier transform (FT) from U, V to l, m is applied to the U, V, W -data we need to deal with the phase term expression given by (3.21). The first term is removed by the fringe stopping leaving

$$\begin{aligned} \varphi_s = 2\pi [& (l_s U' + m_s V') - (l_s^2 + m_s^2) W / 2n_0 \\ & - (l_0 l_s + m_0 m_s)^2 W / 2n_0^3] \end{aligned} \quad [\text{rad}] \quad (3.22)$$

where we introduced $U' = (U - W l_0/n_0)$ and $V' = (V - W m_0/n_0)$. The first term between parentheses in (3.22) contains the 2-D Fourier kernel and gives an image in the l, m -plane that is centred on l_0, m_0 and uses the shifted coordinates l_s and m_s and values for U' and V' that are corrected for a baseline tilt relative to the U, V -plane as caused by W -projection.

The second and third term between parentheses contribute to the W-term (3.12) with terms quadratic in l_s and m_s and a cross product term ($l_s m_s l_0 m_0$) that is quadratic on the “diagonal” with $l_s = m_s$ and smaller elsewhere.

Without correction for the W-contributions we obtain an image where sources at larger distance from the centre will be distorted by the phase distortion $\Delta\varphi_s$ which is dominated by terms that are quadratic with distance from the field centre

$$\Delta\varphi_s = \pi [-(l_s^2 + m_s^2) - (l_s l_0/n_0 + m_s m_0/n_0)^2] W / n_0 \quad [\text{rad}] \quad (3.23)$$

The phase distortion in the visibility of each baseline produces a distorted side lobe pattern around each source in the field after a 2-D FT. For a source at larger distance from the field centre the phase distortion increases quadratic with increasing distance from the centre of the Fourier image. These distorted side lobes cause additional noise when only the nominal side lobes are removed by subtraction in an image of a scaled point spread function (psf) determined by the U',V'-distribution of the 2-D IFT. An important aspect is that the phase distortions do not have a random value over all baselines, but have a certain structure since the stations follow Earth curvature. This means that a station at larger distance from the centre of the array creates a larger phase distortion on all its baselines to the stations near the centre of the array. In effect the phase distortion increases quadratic with this distance due to Earth curvature.

We now look at the effective FoV for two special cases, (i) W-axis towards centre of the field, i.e. $l_0 = m_0 = 0$ which allows ignoring the second term between parentheses in (3.23) and (ii) $l_0 = 0$ and $m_0 > m_s/n_0$ where the second term dominates.

The first case is typical for a synthesis array where W varies during an observation and a typical value is half the maximum baseline in wavelengths. For a small angular distance $\Delta\theta_r$ from the field centre we have in the first case $n_0 \sim 1$ and we find a phase distortion determined by $l_s^2 + m_s^2 = \Delta\theta_r^2$ that is given by

$$\delta\varphi_s \sim \pi W \Delta\theta_r^2 \quad [\text{rad}] \quad (3.24)$$

If we tolerate a maximum phase distortion $\delta\varphi_s = \pi^{-1}$ the radius of the minimally distorted Field-of-View (FoV) is defined by

$$\Delta\theta_r = \pi^{-1} W^{-1/2} \quad [\text{rad}] \quad (3.24a)$$

A maximum phase distortion $\pi^{-1} \sim 0.3$ rad to simplify the formula seems somewhat arbitrary but gives an acceptable degradation of the intensity of a source at the edge of the FoV, which can be illustrated as follows.

If we assume that the phase distortions in the visibility samples along a U,V-track are uniformly distributed between $-\delta\phi$ to $+\delta\phi$ then the intensity is degraded by a factor $\langle \cos(\delta\phi) \rangle = \text{sinc}(\delta\phi) \sim (1 - \delta\phi^2/6)$ and leads only to a degradation of 1.7 % (1-0.983) for $\delta\phi = \pi^{-1}$, while the visibility phase makes a saw-tooth pattern with zero average. This degradation on baselines with the largest non-planarity has a magnitude comparable to that caused by integration time smearing and bandwidth integration on the longest baselines as will be discussed in section 3.2.

The maximum degradation for a limited set of baselines applies only to objects at the edge of the FoV and is considered acceptable since these objects are already reduced by more than 50% by the station beam. For the coming sections we stick to the imaging FoV definition given by (3.24a) that defines a 1.7 % loss in sensitivity in the visibility of a source at the FoV radius from the field centre on the baseline with largest non-planarity

The second case is typical for a quasi-planar array with height deviation H from the reference plane and for $m_0 = \sin \theta_0$ we get $m_s = \Delta\theta_0 n_0$ since for $l_0 = 0$ we have $n_0 = \cos \theta_0$ and we get

$$\delta\varphi_s \sim \pi \Delta\theta_0^2 m_0^2 n_0^{-1} H \lambda^{-1} \quad [\text{rad}] \quad (3.25)$$

Again tolerating a maximum phase error $\delta\varphi_s = \pi^{-1}$ the maximum Field-of-View (FoV) extent in elevation from the centre of the image is defined by

$$\Delta\theta_0 = \pi^{-1} \lambda^{1/2} H^{-1/2} n_0^{1/2} m_0^{-1} \quad [\text{rad}] \quad (3.25a)$$

Although the actual distribution of the phase errors by the various H terms in a non-planar array on a curved Earth depends on the actual array configuration we assume that the degradation is of the same 2 % order as for baselines with stations at a large distance from the core.

Apart from the reduced flux in a point source imaged away from the field centre there are also enhanced side lobes and their impact will be eliminated by subtracting the strongest sources from the visibility data as will be further discussed in section 3.1.11. Although a certain tolerance on side lobe distortion could be acceptable for a single observation, in practice many observations could be averaged to improve the sensitivity. Systematic phase errors in the image forming would then determine the observed noise floor in the averaged observation and will be discussed in chapter 5.

3.1.9 FoV for Intrinsic and extrinsic non-planarity

Intrinsic deviations from planarity are mainly caused by Earth curvature. Their typical value is then given by

$$H = L^2 / 2 R_E \quad (3.26)$$

Where L is the distance of a station from the centre of the array and R_E is the Earth radius $\sim 6,371$ km. An impression for the LOFAR situation is given in table 3.1.

Table 3.1. Earth Curvature H for LOFAR stations at distances L from the array centre

L km	1	2	4	8	15	30	80	200	400	600
H	0.08 m	0.31 m	1.26 m	5.0 m	18 m	71 m	500 m	3.1 km	12 km	28 km

Interestingly there are 4 LOFAR stations at a distance of ~ 600 km from the centre, which means that they all four lie in the plane of a spherical cap with the centre of LOFAR at the top. All baselines between stations at the edge of the cap are parallel to the tangent plane at the centre of LOFAR. The baselines between these stations could become as large as 1200 km, but they are all co-planar to the short baselines in the horizontal plane at the centre of the array. All baselines between core stations and remote stations at 600 km distance are not co-planar.

LOFAR has its largest FoV when observing at 50 MHz with the compact LBA station configuration that has a diameter $D = 32$ m. This results for wavelength $\lambda = 6$ m into a typical beam radius at half power of $0.6 \lambda / D = 0.11$ assuming a taper that reduces the side lobe pattern. This half width is valid for a phased array with its beam pointed towards Zenith and is larger in elevation direction at lower elevations. If we make a snapshot image out to $\Delta\theta = 0.11$ rad from the beam centre at zenith angle $\theta_0 = 45^\circ$ we find according to (3.25) $H_{\max} \sim 72$ m which allows simple 2-D Fourier imaging with stations out to 30 km from the centre of the array. However, doubling the field radius to 0.24 reduces H by a factor 4 but halves L to 15 km. Baselines with stations that are further away will need some form of correction as will be discussed in section 3.4.

Extrinsic deviations from planarity arise when we define a reference plane for a 2-D Fourier transform that makes a tilt θ_0 with the plane of a planar array. If two stations have a separation B we could reach a maximum extrinsic non-planarity $B \sin(\theta_0)$ for the baseline between the two stations.

If we use for the same LOFAR array a 2-D Fourier inversion in a coordinate system with W-axis towards the field centre we get a projection $W = B_{\max} \sin(\theta_0)$, which allows for a FoV at zenith angle $\theta_0 = 45^\circ$ and a radius 0.11 a maximum baseline $B_{\max} \sim 72$ m. Halving the FoV radius would allow a four times larger maximum baseline of 288 m. Apparently, we need a number of smaller fields to cover the full station beam at full resolution as will be further discussed in section 3.4.

3.1.10 Synthesis imaging with a single 2-D Fourier inversion

Earth rotation causes a continuous change of the angle between baseline and source direction and results in a 3-D distribution of baselines, which can be inverted to an image by a 3-D Fourier transform as discussed in subsection 3.1.2. This 3-D transform uses a large number of 2-D Fourier transforms and it would be more attractive to use a single 2-D Fourier transform based on (3.22). In that case we need to deal with phase distortion term (3.23), which leaves us two options

- (i) limit l_s and m_s i.e. limit the maximum extent of the imaged field,
- (ii) limit W , i.e. limit the non-planarity, intrinsic as well as extrinsic.

Conventional synthesis imaging follows (i) by choosing a coordinate system with W-axis towards the centre of the source field of interest. The advantage is that a single l_s, m_s -coordinate system is obtained with a single point spread function, which simplifies further processing of the image. The disadvantage is that due to the large extrinsic non-planarity, which includes the intrinsic contributions, only a small FoV can be imaged as determined by (3.24). In practice this method works remarkably well for small arrays and small fields as provided by telescopes of the 25 m class that operate at high frequencies. However, for LOFAR operating at low frequencies and long baselines we need either many small fields or corrections that will be discussed in section 3.4.

Alternative method (ii), sometimes called snapshot imaging, limits W to just intrinsic non-planarity by Earth curvature. A reference plane is chosen that minimizes the height differences with the stations, such as a plane perpendicular to local Zenith at the centre of a symmetric array. The advantage is a much larger FoV, in principle as large as a hemisphere for a true planar array. The disadvantage is that snapshot images need to be made with limited duration, since the second term between parentheses in (3.23) contains l_0 and m_0 that move relative to the coordinate system that is fixed to the Earth. This approach needs therefore a number of 2-D Fourier transforms to handle a long synthesis observation. In section 3.5 we will show that the maximum duration of a snapshot image is matched to the timescale of the changing beam shape, which is of order 10 min as will be analysed in section 3.6. This allows for LOFAR a much smaller set of 2-D Fourier transforms than the 3-D approach and allows correction for the beam foreshortening per image.

3.1.11 Point Spread Function

The Point Spread Function (psf) is the impulse response function of the instrument and is for each direction on the sky given by the image produced by a point source at that location.

In case the psf is shift-invariant, as for a true Fourier image, it needs to be evaluated only for a source at the centre of a 2-D FT image and is then valid for all positions in the image field. A 2-D Fourier image with a non-planar array has distorted point sources, of which the distortion increases with distance from the field centre. Convolution with the nominal psf gives a distorted side lobe pattern as function of source position, but the pattern itself is position invariant.

A fast Fourier transform (FFT) based digital spectral filter bank provides a set of subbands that form the input for the station beam forming and station correlation processing. In the spectral domain we need to realize that each channel formed by the FFT also contains signals picked up by the side lobes. It is therefore important that a convolution filter in the time domain provides a proper filter in the spectral domain that attenuates these side lobes. Effective attenuation as realized in a poly-phase filter bank can be obtained by simple convolution filters since data sampling is already available on a regular time grid [Alliot, 2002], but artefacts appear for signals that violate the requirement of stationarity, as is the case for some types of interfering signals.

In section 3.4 we will discuss an approach for spatial convolution that on the one hand effectively limits the FoV for the FFT, but on the other hand allows performing corrections that make a point source appear as a point source in a limited 2-D FFT image.

3.1.12 Combining direct and model based inversion to handle non-planarity

The simple 2-D Fourier inversion shown in (3.14) provides a proper image if a few conditions are fulfilled such as planarity of the array as well as having identical station beams. In practice, these requirements are only partially fulfilled due to e.g. Earth curvature and due to particular design considerations. This could be the rotation of the LOFAR station beams in an attempt to reduce the side lobe effects in the average beam pattern of all phase array stations together. The inversion problem is then practically handled by turning it into a problem whereby a model of the observed reality is adopted for which parameters have to be estimated. These approaches are computational intensive, but turn out to become possible for real life astronomical array systems of limited size [Wijnholds, 2010].

Practical astronomical imaging packages use a hybrid method that combines (i) solving for complex gain parameters using the few strongest sources in the visibility

data, (ii) subtracting these sources from the visibility data that would otherwise mask weaker sources, and (iii) approximate Fourier imaging to identify a next set of weaker sources. In a few iteration steps the calibration is improved using the additional sources and the few strongest sources are subtracted accurately from the visibility data using the improved calibration parameters. Assuming that all sources in the field need the same final calibration parameters more sources are subtracted in subsequent iteration steps without improving the calibration. The process is ended when the nominal side lobes of all remaining point sources are below a predefined fraction of the expected thermal noise in the image. We are finally left with a set of subtracted model sources and a residual image with all weaker sources and three types of noise contributions. We have side lobe noise by all sources that are not subtracted and residual side lobe noise by all sources that are subtracted but needed slightly different calibration parameters. The third contribution is the nominal thermal noise and in chapter 5 we will express the two so called non-thermal noise contributions by side lobes as fraction of the thermal noise.

The described procedure assumed that initial instrumental calibration parameters such as the complex gain and receiver pass band per station are derived from separate observations where the visibility data is dominated by a single strong source in the main beam of the stations. For a synthesis imaging observation we need a correction per station for varying phase and gain induced by electronics, troposphere and ionosphere, and for amplitude due to beam shape variation during that observation. This latter automatically includes the effects for differences between individual station beams.

Differences between the –voltage- main beam g_{ik} of the stations i for direction k are small for almost equal stations and a single –power- main beam pattern $g_k^p = g_k g_k^*$ that is some average over all station beams is adequate for approximate 2-D Fourier inversion. The differences between the station beams create differences in the effective gain of the interferometer signals for each direction. On average, we get for the synthesized beam of the array a psf with nominal unity gain for the field centre, but its side lobe pattern will have an additional error pattern that is different for each direction within the average main beam of all stations. These station beam differences play only a role for the weaker sources that have not been subtracted in the visibility domain using proper gain factors for each station.

3.1.13 Summary, Conclusions and main Result

From the basic equation for the response of an interferometer we derived an exact 3-D imaging equation that is suitable for Fourier inversion provided that certain conditions are met. We analysed how well these conditions can be met by the actual LOFAR synthesis array and station configuration and which approximate Fourier inversions could be useful to obtain wide field imaging that will be computationally affordable.

The first three results are

- Fourier inversion assumes that all stations have an identical station beam pattern.
 - Differences between station beams result in direction and baseline dependent amplitude errors that lead to distortion of point sources depending on their location within the station beam.
 - Such errors are calibrated just as the phase errors that will be further discussed in chapter 4.
- 3-D Fourier inversion chooses the W-axis of the reference coordinate system towards the source that is being tracked and handles all non-coplanar baselines of a full synthesis observation.
 - There is however a processing penalty for Fourier inversion defined as the ratio of generated 3-D image points over the number of actually used image points on the l, m, n -unit-sphere.
 - This Fourier processing penalty is the consequence of the large number of 2-D FFTs that need to be performed for a number of n -values, which is determined by the length of the baseline projection on the W-axis.
 - For a ~ 6 h synthesis this W-range could reach a value $\sim B_m/2\lambda$, where B_m is the maximum baseline and λ the wavelength.
 - For LOFAR with baselines up to 120 km and a station beam with a FWHM ~ 0.24 rad at 50 MHz the number of 2-D Fourier planes can be larger than ~ 400 , even for a single snapshot image with an array that is planar itself. This makes the method impractical to use for LOFAR.
- 2-D Fourier imaging leaves only small non-planarity dependent phase errors for sources at the edge of the field, if the field is sufficiently small.
 - Although the psf of the 2-D Fourier transform is position independent the distorted point sources have a distorted side lobe pattern depending on the position of the source.
 - However, subtraction of sources from the visibility data including the position dependent imaging error per baseline removes the artefacts from the image completely.
 - In contrast, calibration of ionosphere induced phase distortions leaves residual phase errors; these errors are random and average out when independent ionosphere intervals are combined, but leave additional noise in an image.

Our main result (3.22) is a second order expansion of the term $(1 - l^2 - m^2)^{1/2} W$ in the 3-D Fourier kernel that gives apparent phase errors in the visibilities when a 2-D

Fourier inversion is attempted for a non-planar array. The expansion includes the contribution of the shift of the image centre to an arbitrary position (l_0, m_0, n_0) on the unit sphere defined for the arbitrarily chosen Cartesian U,V,W- coordinate system that describes the baselines of the interferometer array. This effect gives a major limitation to the maximum duration of a synthesised snapshot image with W-axis towards local Zenith of the array centre to be discussed in section 3.5.

- Conventional synthesis chooses the W-axis towards the source field that is tracked, such that $(l_0, m_0, n_0) = (0, 0, 1)$. Inversion with a 2-D Fourier transform then gives distorted images since the visibilities of a source show a dominant phase error proportional to $(l^2 + m^2) W$.
 - Just as for the 3-D case the W-contribution is so called extrinsic non-planarity, which is the consequence of a particular choice of the coordinate system that ignores the intrinsic planarity of the array.
 - Equation (3.24) gives a first order estimate for a FoV where objects suffer mainly from limited phase distortion by quadratic terms in l and m
 - LOFAR with a maximum station beam diameter in Zenith of $\sim 12.5^\circ$ FWHM at 50 MHz has a maximum phase error of 0.3 rad at the edge of a limited field with radius of 3.1° supporting a longest baseline of only 288 m.
- An alternative 2-D snapshot Fourier inversion is proposed for a quasi-planar array where a large field is allowed since the W-terms are limited by choosing a coordinate system with W-axis perpendicular to the best fit plane of the array.
 - For a W-axis towards local Zenith of the centre of a synthesis array, the non-planarity of the array is limited to intrinsic values that are mainly determined by Earth curvature, which are much smaller than the maximum baseline.
 - First and second order terms in the expansion of $(1 - l^2 - m^2)^{1/2}$ cause phase contributions in 3-D visibility data that have to be corrected for a small field 2-D Fourier transform with differential coordinates l_s, m_s . These phase contributions are proportional to first and second order terms of the a fringe shift of the centre of the source field to (l_0, m_0, n_0) .
 - After a fringe shift correction that tracks a sky source we find non-planarity phase errors that are enhanced by a factor $n_0^{-1} = \sin(\text{elevation})^{-1}$, which gives a practical limitation at very low elevations. LOFAR will preferably not observe for imaging purposes at too low elevations in view of its elongated station beam and worse ionosphere effects.

- Instead of projection along the W-axis, as for the conventional case, we find U,V-coordinates for the shifted field with l_s, m_s -coordinates by projection along the source direction; these projections are the result of the first order shift contributions.
- Apart from 2nd and higher order terms there is an important cross term with $(l_s m_s l_0 m_0)$ that can be reduced effectively by limiting l_0 to a small tracking range δl_t if the l-axis of the snapshot frame is properly chosen as will be discussed in section 3.5. Instead of short duration snapshots we then form longer synthesized snapshots.
- In contrast to even order terms that give a constant phase error the odd order phase term with δl averages to zero but causes an amplitude degradation.
- This degradation can be compared with degradation by bandwidth and integration intervals that will be discussed in section 3.2.
- The full FoV of the station beam out to a radius of 6.3° of a 32 m LOFAR station at 50 MHz is at 45° elevation still properly imaged with 2-D snapshots for stations out to 30 km from the centre of the array if a maximum phase deviation of π^{-1} is tolerated from second order terms.
- Depending on the allowed degradation by tracking time of each snapshot a number of synthesized snapshots are required that need combining with appropriate interpolation; this is a different interpolation than needed in the 3D case.
- Such a number is just needed to correct every snapshot image for shape changes in the average polarized station beam.

The main result can be summarized as follows

- Fringe shifted visibility data of a quasi-planar array contain phase terms proportional to the non-planarity and fringe shift
- 2-D Fourier imaging in the plane of the array centred at the shifted position needs as first order correction projection of the baselines on the plane of the array from the direction of the field centre.
- The second and higher order phase terms limit the FoV of the shifted Fourier image as function of non-planarity and fringe shift.

3.2 Decorrelation by averaging in frequency and time domain

In this section we discuss the well-known effects of integration over frequency and time by cross-correlation as will be used in further analysis. Averaging over frequency leads for a planar phased array in the horizontal plane to the so called beam squint effect where the shape of the array beam distorts as function of its distance from the Zenith direction.

The product of baseline vector \mathbf{U}_{ij} of an array with source direction \mathbf{l}_k changes due to Earth rotation and causes a more rapid phase change for longer baselines (see (3.4)) and for larger distance of \mathbf{l}_k from the direction of a celestial pole. For finite integration time, this leads therefore to degradation of the correlated visibility of an object as function of baseline length and distance from the pole.

A detailed analysis [chapter 18, Taylor, 1999] will be summarized for the worst case situation that is relevant to define appropriate channel bandwidth and appropriate integration time for the cross-correlation processing to allow sufficient imaging quality over a wide image field.

3.2.1 Tolerated amplitude degradation

The problem is analysed by looking at the phase term in the exponent of (3.4) where the baseline vector in wavelength units \mathbf{U}_{ij} rotates with respect to source direction \mathbf{l}_k due to Earth rotation and changes length proportional to frequency. More precisely we look at the phase $\Delta\varphi$ of a source at position \mathbf{l}_k relative to the phase of a source at position \mathbf{l}_0 for which the appropriate fringe tracking corrections are done for baseline vector \mathbf{B} [m] at wavelength λ [m].

$$\Delta\varphi = 2\pi \mathbf{B} \cdot (\mathbf{l}_k - \mathbf{l}_0) / \lambda \quad [\text{rad}] \quad (3.27)$$

The effect of the fringe tracking operation is that it reduces the phase rate of $\Delta\varphi$ and allows limited amplitude degradation of the complex visibility $P_k \exp(-i \Delta\varphi)$ as function of integration time τ of the correlation for objects at some distance $\Delta\mathbf{l} = \mathbf{l}_k - \mathbf{l}_0$ from the tracking position.

We simplify the analysis by replacing $\mathbf{B} \cdot (\mathbf{l}_k - \mathbf{l}_0)$ in (3.27) by $B \Delta\mathbf{l} \cos(\chi)$, with vector length's B and $\Delta\mathbf{l}$ and angle χ between the vectors. For an object at a given distance $\Delta\mathbf{l}$ from the fringe stopping position, which is normally chosen equal to the pointing position of the stations in the synthesis array, we have then a phase

$$\Delta\varphi = 2\pi B \Delta\mathbf{l} \cos(\chi) / \lambda \quad [\text{rad}] \quad (3.27a)$$

For a phase $\Delta\varphi$ that varies linearly from $\Delta\varphi - \delta\varphi/2$ to $\Delta\varphi + \delta\varphi/2$ as function of one of its parameters we get an averaged visibility $\langle c_{ijk} \rangle$ given by

$$\langle c_{ijk} \rangle = c_{ijk} \text{sinc}(\delta\varphi/2) \exp(-i \Delta\varphi) \quad [\text{rad}] \quad (3.28)$$

Apparently the visibility amplitude is degraded by the sinc factor that for small total phase change $\delta\varphi$ can be approximated by $(1 - \delta\varphi^2/24)$ and we call the term $\delta\varphi^2/24$ the degradation.

A maximum phase deviation $\delta\varphi_m/2$ on the longest baseline means proportionally smaller phase deviations on shorter baselines and quadratically less degradation. When the imaging process averages the visibilities over all baselines we get a small broadening of the point spread function by the change in its effective taper as well as a small degradation of the effective signal to noise ratio. This degradation is however proportional to the square of the distance from the centre of the field. A more detailed analysis [chapter 18, Taylor, 1999] takes into account the averaging over a full synthesis image with varying $\delta\varphi_m$. To stay consistent with the other FoV degradation we use $\delta\varphi_m/2 = \pi^{-1}$ on the longest baseline as the maximum, for an object at half a beam width distance from the centre of a station beam. This means only 1.7 % worst case degradation on the longest baseline for objects at the edge of the FoV defined by the half power level of the station beam and less than 1% for the average of all baselines, depending on their relative weight in an image.

3.2.2 Time averaging

We need to evaluate (3.27) as function of integration time where χ is a function of vectors \mathbf{B} , \mathbf{l} and \mathbf{l}_0 . We simplify to a worst case situation with an interferometer located at an Earth pole that tracks the pole. In that case

$$\delta\varphi_\tau = 2\pi \omega \tau B \Delta l \sin(\chi) / \lambda \quad (3.29)$$

Where ω is the time derivative of χ which then equals the Earth rotation of $72.7 \cdot 10^{-6}$ radians per second and τ is the integration interval. A station with diameter D and parabolic aperture tapering has a half power beam width of $1.28 \lambda/D$. For a source at half power we take $\Delta l = 0.6 \lambda/D$, which leads for $\chi = \pi/2$ to a maximum visibility degradation of 1.7 % on baseline B to $\delta\varphi_\tau/2 < \pi^{-1}$ and we find

$$\tau < 2323 D/B \quad [\text{s}] \quad (3.30)$$

Interestingly, the maximum integration time is independent of frequency and we evaluate τ for a number of representative values of D and B for LOFAR in the following table 3.1.

Table 3.1. Integration time τ [s] as function of baselines and LOFAR station diameter giving < 1.7 % degradation

D m	B) ¹	2 km	6 km	20 km	60 km	90 km	300 km	600 km	1200 km
28) ²	HBA _C	32	11						
32) ³	LBA _S	37	12	3.7	4.1	2.8	0.25	0.12	0.06
40 m	HBA _R	46	15	4.6	1.5	1.0			
56 m	HBA _E					1.5	0.40	0.20	0.10
68 m	LBA _E					1.8	0.52	0.26	0.13
81 m	LBA	94	31	9.4	3.1	2.1			

)¹ Index of station type for Core, Small configuration, Remote, and European

)² Equivalent HBA diameter of circular area by total number of elements each providing $(5.14/4)^2 \text{ m}^2$

)³ LBA diameter given by longest separation between antenna elements

3.2.3 Frequency averaging

We evaluate (3.27) for a phase change $\delta\phi_v$ that occurs when the wavelength is changed by $\delta\lambda$ from $\lambda - \delta\lambda/2$ to $\lambda + \delta\lambda/2$ for $\chi = 0$ at separation $\Delta l = 0.6 \lambda/D$ from the field centre and find

$$\delta\phi_v = 2\pi 0.6 \delta\lambda/\lambda \quad B/D \quad (3.31)$$

We tolerate again a degradation of 1.7 % so $\delta\phi_v/2 < \pi^{-1}$ and we find

$$\delta v/v = \delta\lambda/\lambda < 0.168 D/B \quad (3.32)$$

For $\chi = 0$ both the diameter D of a phased array station and the baseline length B have the same elevation dependent foreshortening which leads to phase error $\delta\phi_v$. This error is independent of elevation although the station beam broadens in angular elevation extent at lower elevation.

Interestingly we see that $\delta v/v$ and τ have the same dependence on D/B which allows expressing δv as function of v and τ according to

$$\delta v = 0.0723 \tau v \quad [\text{kHz with } v \text{ in MHz}] \quad (3.33)$$

This allows using table 3.1 also to evaluate $\delta\nu$ for continuum imaging according to (3.33), while line observations might need narrower spectral channels. For convenience we give in table 3.2 the maximum channel bandwidth as function of station diameter for representative frequencies and baselines.

Table 3.2. Maximum channel bandwidth $\delta\nu$ [kHz] as function of frequency, station diameter D and baseline B.

Freq MHz	B D	2 km	6 km	20 km	60 km	90 km	300 km	600 km	1200 km
35	68 m					4.4	1.3	0.7	0.3
	81 m	238	79	24	8	5.3			
70	32 m	188	63	19	6.3	4.2			
	68 m					9	2.7	1.3	0.7
	81 m	119	40	12	4	2.6			
140	28 m	329	110						
	40 m	470	157	47	16	10			
	56 m					15	4.4	2.2	1.1

Maximum degradation 1.7 % for objects at half power of station beam pointing at Zenith.

If the required field size is limited, as in facet imaging, we need to replace the station diameter D by some equivalent larger diameter and take properly into account the FoV definition. The FoV is based on a maximum degradation of 1.7 % in the visibility on the longest baseline of a source at half power of a circular station beam (i.e. when pointed towards Zenith). If a higher degradation is accepted, a larger bandwidth and longer integration time could be used, which greatly reduces processing as will be shown section 3.7.

3.2.4 Effects of the sinc shaped degradation function

The effect of integration of a phasor over a small phase range is a small amplitude degradation and has already been discussed in subsection 3.2.1. For a given channel bandwidth and integration time the degradation scales with the square of the baseline and with the square of the distance of an object from the field centre. This degradation results in a lower weight of the visibility at longer baselines of a point source at larger distance from the field centre. This reduced weight leads not only to reduced intensity but also to effective broadening of these point sources, which leaves residuals if a scaled nominal point spread function is subtracted. A more detailed analysis of these effects that also include the shape of the profile that is

used for integration over a time and a bandwidth interval is given in chapter 18 of [Taylor, 1999].

For larger distances from the fringe stopping centre the phase range increases proportionally and when it exceeds 2π the sinc function, valid for uniform integration intervals, gets a periodic behaviour with slowly decreasing peak values. This means that distant objects are not well attenuated with distance, and suggests to use an integration function that gives steeper decay than the block integration.

As will be discussed in section 3.3.2 smoothing of the sinc function could in principle be realized by the convolving regridding process. In case that the regridding uses only few surrounding visibility samples, it is important that each sample already has high attenuation for sources outside the FoV. For LOFAR with its FX correlation system implemented on a High Performance Computing (HPC) system it is in principle possible to implement integration intervals with a profile that gives high attenuation for distant objects, especially at the long baselines.

3.2.5 Correlation and post correlation processing impact

The worst-case values provided by (3.30) and (3.33) are useful to specify the correlation process and define a minimum correlation output sample rate for the longest baselines for a given total bandwidth per station. As a result, the data output rate of the correlation processing is at least proportional to the FoV expressed in resolution elements. Sources at half power level of the station beam will get at most 1.7 % degradation on the longest baselines. However, for sources that are already 50% attenuated or more by the station beam, an additional sensitivity loss at the edge of the FoV of less than a percent (for all baselines together) seems not very critical. Apart from minor source broadening, discussed above, we suffer from degraded survey sensitivity when sources at the edge of the field are attenuated. This loss in survey sensitivity could be compensated by sampling the sky with a grid of station beams, which reduces the total duration of a given survey area.

This analysis shows that the survey performance of a synthesis array is degraded by the limited output data rate of an FX correlation platform. In practice, spectral and temporal resolutions define an output data rate that is much smaller than the input data rate determined by number of stations and bandwidth per station. As a result, the cost of an FX correlation platform is almost independent of output sample rate, but proportional to station bandwidth and to the total number of baselines between all stations. In contrast, the cost of a post correlation platform that needs to keep up correcting all the visibility samples in real time is proportional to the output sample rate of the correlation platform. This shows that the choice of the maximum degradation on the longest baselines critically determines the marginal sensitivity cost of the post correlation processing, which could simply be balanced against the

marginal sensitivity cost of the array that brings in the sensitivity and the survey speed [Bregman, 2010].

3.3 Fast Fourier Transform imaging and filtering by Convolution

Coherencies c_{ij}^m measured with a correlation interferometer contain according to (3.7) not only the signals of all sources that are visible by the antenna stations, but contain also instrumental, averaging and sampling effects expressed by (3.5).

The sampling theorem states that if a -noise like- signal is band-limited, it can be completely represented by a set of samples spaced by the reciprocal of twice the bandwidth (conventional Nyquist sampling).

In case of an interferometer where the elements are phased array stations, the field is first limited by the beam of the antenna element in a station and secondly by the array beam of the station. This process limits the low spatial frequency components, while the longest baseline limits the high spatial frequencies. This means according to the convolution theorem for Fourier transforms (FT) that the samples c_{ij}^m in (3.3) are convolved with the FT of the -power- beam pattern $g^p = g_k g_k^*$ which equals the aperture illumination pattern of the stations and therefore indeed has a finite extent.

Because the sampling function S_{ij} in (3.7) is a multiplicative factor, the image which results from 3-D or 2-D inverse Fourier transformation (IFT) of measured visibilities is a convolution of the true sky with the Fourier inverse of that sampling function. Poor sampling of the U,V,W-space or U,V-plane therefore introduces side lobes and grating lobes as determined by the distribution of receptors in the correlation array with respect to the sky.

Practical implementations of the Fourier inversion use the fast Fourier transform (FFT) for which the required processing capacity scales with $N_p \log(N_p)$ instead of $N_p N_{bs}$ as for the FT, where N_p is the number of image pixels and N_{bs} the total number of baseline samples that is used as input. Since $N_{bs} \gg \log(N_p)$ it is computationally attractive to use the FFT instead of an FT.

Unfortunately the FFT assumes its input samples on a rectangular grid, and hence a convolution operation is required before resampling to that grid. The convolution kernel has N_k pixels requiring additional processing capacity proportional to $N_k N_{bs}$.

In this section we analyse the effects of sampling, convolution, and resampling.

3.3.1 Resampling convolution of observed interferometer data

The following section explains the various steps in sampling and convolution filtering that are well known in principle but are now placed in the relevant context that is not readily found.

We start our discussion with the planar version (3.11) of the measured visibilities and apply a resampling convolution function $C(\mathbf{U}) = C(U, V, W=0) = C(U, V)$ that adds weighted values for all observed data c_{ij}^m at each point of a rectangular grid (p, q) within the kernel extent around each observed data point according to

$$c_{pq}^r = \sum_{ij} C((U_{pq} - U_{ij}), (V_{pq} - V_{ij})) c_{ij}^m \quad (3.34)$$

Although the convolution is only calculated for the grid points the operation is mathematically described by a convolution with function C_{pq} followed by a sampling operation [chapter 10.2, Thompson, 2004]. The resulting resampled data c_{pq}^r are then described by

$$c_{pq}^r = S_{pq}^r (C_{pq} * c_{ij}^m) \quad (3.35)$$

where $*$ indicates the convolution operation and S_{pq}^r is the resampling function to a rectangular grid applied after that convolution.

Closer inspection of (3.11) shows that c_{ij}^m is the product of the interferometer sampling function S_{ij} and the 2-D Fourier transform (FT) of our sky image $I_k^s = dl_k dm_k n_k^{-1} P_k$ multiplied with power beam P_k and with W -term G_{kij} . The whole process can therefore be described by

$$c_{pq}^r = S_{pq}^r (C_{pq} * (S_{ij} F(g^p G_{ij} I^s))) \quad (3.35a)$$

In which $F(\)$ represents the FT. In this formulation we dropped the position index k of the sky image I^s of the power beam pattern g^p and of the W -term G_{ij} (3.12) since this index is eliminated by the FT. An important constraint is that the convolution kernel C_{pq} is at least one grid cell wide to provide an effective beam pattern in the image domain.

If we take the inverse Fourier transformation F^{-1} of (3.35a) and apply the convolution theorem to the right hand side we get

$$F^{-1}(c_{pq}^r) = F^{-1}(S_{pq}^r) * F^{-1}(C_{pq} * (S_{ij} F(g^p G_{ij} I^s)))$$

Executing the inverse Fourier transforms and again applying the convolution theorem we obtain the regridded image $I_{lm}^r = F^{-1}(c_{pq}^r)$ given by

$$I_{lm}^r = \underline{S}_{lm}^r * (\underline{C}_{lm} F^{-1}(S_{ij} F(g^p G_{ij} I^s))) \quad (3.36)$$

The inverse FT of C_{pq} is an additional beam pattern \underline{C}_{lm} in the image domain. Executing the inverse Fourier transform and applying the convolution theorem in the right hand side gives

$$I'_{lm} = \underline{S}'_{lm} * (\underline{C}_{lm} (\underline{S}_{lm} * (g^p G_{ij} I^s))) \quad (3.37)$$

This equation shows that the output I'_{lm} of the FFT image contains a modified sky image I^s that is first multiplied with the station power beam pattern g^p and with complex W -term G_{ij} and then convolved with \underline{S}_{lm} the Fourier transform of the interferometer sampling function S_{ij} . This convolution creates side lobes and grating responses in the field of interest that emanate from sources outside that field and can only be eliminated by subtracting the sources from the correlation data using an accurate model that includes the full G_{ijk} .

The convolved result is attenuated by an additional beam pattern \underline{C}_{lm} that is the Fourier inverse of the convolution function C_{pq} . Finally the resulting image is convolved with the Fourier transform of the resampling grid that replicates the image field over the sky and therefore contains aliased images of the rest of the sky.

Fortunately these replicated sky images are attenuated by the additional beam pattern \underline{C}_{lm} . This requires that the beam \underline{C}_{lm} needs a steep but smooth decay outside the area of interest to reduce these contributions. The sampling function S'_{pq} has a finite extent, i.e. is multiplied with a top hat, so its Fourier transform \underline{S}'_{lm} (that is also a grid of δ -functions) is convolved with the Fourier transform of the hat function.

A planar array has a hemispheric FoV when a 2-D FT is used, and we have proper images all over the sky that might appear attenuated and aliased in an image made with only a small FFT field. Partly this attenuation is caused by the beam \underline{C}_{lm} as a result of the convolution, and partly by integration time and bandwidth smearing as discussed in subsection 3.2.4. The latter two amplitude effects cause also object distortions depending on the distance of an object from the fringe tracking position, since different baselines have different attenuation depending on their length and orientation. These distortions are different from distortions by non-planarity.

3.3.2 Distortion correction by convolution

In section 3.1.6 it was explained that data taken with a set of interferometers that are not in a plane show phase deviations related to the projection of the baselines on the W -axis which leads to distortions of objects in a 2-D Fourier image. Our actual measurements c^m_{ij} contain these deviations that can in principle be corrected by a complex convolution operation. An important distortion is the G_{ijk} term that appears in (3.11) which could be corrected by extending the resampling convolution C_{pq} with an imaginary term as will be discussed in section 3.4. Retaining the real part \underline{C}^R_{lm} of \underline{C}_{lm} gives

$$I'_{lm} = \underline{S}_{lm} * (\underline{C}_{lm}^R (\underline{S}_{lm} * (g^p I^s))) \quad (3.37a)$$

The corrections will be accurate for objects within the additional beam pattern \underline{C}_{lm} and only partial for objects further away from the fringe stop position. However the point spread function (psf) \underline{S}_{lm} is defined as the 2-D Fourier transform of the interferometer sampling function S_{ij} in the U,V-coordinates of that Fourier transform. The psf is therefore position invariant in its propagation of side lobes from sources in the field. However, sources observed with a non-planar array show phase deviations as function of baseline and distance from the reference position l_0, m_0 as given by (3.23). 2-D Fourier transformation of these visibilities then results in a distorted side lobe pattern, such that a point source has no longer the nominal psf as its side lobe pattern.

It should also be realized that the spiked resampling of the convolved visibility data by S'_{pq} causes a replication of the fields. The result is that signals from adjacent fields appear as aliases as discussed earlier. All objects and all side lobe responses caused by the interferometer sampling S_{ij} that are located outside the FFT field are aliased into the field, but are now progressively attenuated by the additional beam \underline{C}_{lm} (that is aliased as well) before they reach the centre of the image. All sources located inside the FFT field as well as all side lobe responses in the FFT field that are caused by incomplete interferometer sampling S_{ij} of these sources get a limited attenuation by the additional beam. This attenuation is removed by dividing out \underline{C}_{lm} from (3.37) but this enhances the noise at the edges of the FFT field. In practice, only the central part of the FFT field is retained for further processing and we suffer only little noise enhancement in the relevant part of the field.

A good choice for this additional spatial filter \underline{C}_{lm} is the prolate spheroidal function [chapter 7, Taylor, 1999] which is also its own FT [section 10.3, Thompson, 2004]. If the latter is truncated in extent to limit the convolution processing, it has minor impact on the resulting spatial filter [Brouw, 1974].

Unfortunately, the side lobes that result from incomplete sampling S_{ij} of the visibilities of sources outside the FFT field are not attenuated at all by the combination of convolving with C_{pq} before and division by \underline{C}_{lm} after the FFT. Removal of such side lobes requires subtraction of the source response from the visibility data.

Fortunately, sources outside the FFT field are attenuated by bandwidth and integration time smearing of the visibility data without affecting their phase as discussed in subsection 3.2.4. In effect, this smearing extends the size of the interferometer samples in the U,V-domain, which is primarily determined by the aperture size of the station beam. Although the attenuation depends on the broadening of the U,V-sample extent, this broadening depends on the actual U,V-coordinate and the smearing is therefore not a true convolution, but could be considered as a quasi-convolution.

A narrow band snapshot observation by a set of interferometers produces a limited set of visibility samples, each with finite extent as determined by the aperture size of the station beam. When only a small field needs to be transformed by an FFT we get a visibility grid with an increment larger than the size of the visibility samples. The convolution function is typically 7^2 grid points, so every pixel on the visibility grid is filled by the sum of complex visibilities that are actually observed within the support around that grid point. As a result, there is not a contiguous visibility function that is convolved, but every sample on the grid is convolved with a different convolution kernel. Although we expect that the average of all kernels represents the proper one providing the proper additional beam in image space, the actual reduction of signals that originate from objects outside the FFT field might be reduced. A first order estimate for this effect follows the same reasoning as for the side lobe level of the psf of random arrays and then equals $N_{\text{vis}}^{-1/2}$ where N_{vis} is the total number of visibility samples in the snapshot.

3.3.3 Consequences for effective U,V-coverage of line and continuum observations

Missing data between U,V-tracks could prevent proper spatial filtering by 2-D convolution of interferometer data. However, equation (3.37) indicates proper filtering $\underline{C}_{\text{lm}}$ by the convolution of the sampled visibility data with C_{pq} as long as the filter width is larger than the UV-track distance.

Signal amplitude degradation by time and bandwidth averaging as discussed in section 3.2.3 just means lower contribution to the measured average visibility for objects at larger distance from the fringe tracking position. This decorrelation effect could be reduced by using more but narrower spectral channels which allows for “parallel” U,V-tracks from different frequencies. It means in fact more uniform filling of the U,V-plane resulting in lower side lobes and lower grating lobes as will be further discussed in chapter 5. However the processing cost for convolution operations is increased.

For continuum imaging of celestial objects we have intrinsic changes in intensity (non-flat spectrum) and variations over a source with frequency (spectral index varies over the source).

If no U,V-tracks at different frequencies are available as is the case for spectral line objects, we get a line image with a certain side lobe and grating lobe structure. Since the continuum emission has more “parallel” U,V-tracks, it has lower side lobes. After subtraction of the continuum emission, the resulting sources of line emission or absorption show up correctly. However, they have enhanced side lobe structure because of the different side lobe structure of the subtracted continuum emission. A detailed analysis of this subject is outside the scope of this work, but we can already conclude that an array dedicated for line observing such as the core of LOFAR needs a much denser U,V-coverage to reduce the side lobe level of line

sources. Continuum observing allows filling of the U,V-plane by larger bandwidth, which is especially effective in longer baselines, but so called multi-frequency synthesis has its own problems [Rao, 2010].

3.4 Field-of-View extension of 2-D Fourier imaging with non-planar arrays

In this section the computational requirements for increasing the field-of-view with non-planar arrays are considered. It will be shown that extrinsic non-planarity correction of baselines up to 6 km requires for 32 m LOFAR stations at 50 MHz a complex convolution kernel of $\sim 250^2$ pixels to image the full FoV of a station beam. The complex convolution of a single visibility sample then requires about the same processing effort as the correlation. It implies that the processing platform needed for convolution needs about the same processing power to keep up in real time, which is no viable option in practice. Only correcting intrinsic non-planarity could handle easily baselines up to 120 km and would require a more practical kernel size of $\sim 25^2$. However, for longer baselines the kernel size grows due to Earth curvature with the 4th power of the distance between core and furthest station, and requires faceting, in the same way as polyhedron imaging deals with extrinsic non-planarity. A novel Fast Faceting algorithm is presented that makes generation of such a large set of visibility subsets in principle affordable in terms of required processing power. Although datasets are generated for all possible facets, only a much smaller subset with facets centred on relevant objects needs actually be imaged. This approach reduces not only greatly the output of the correlation processing but also reduces the subsequent processing for image forming and deconvolution.

In subsection 3.1.4 it was shown that a set of interferometer measurements could be inverted into a sky brightness image by 2-D Fourier transformation (FT) if some conditions are met. An Earth-bound planar array could then provide an instantaneous hemispheric image, but the 2-D FT image of a non-planar array gets distortions in the point spread function of the objects that are proportional to the non-planarity and increasing with the distance of the objects from the fringe tracking centre as discussed in subsection 3.1.8.

Extending the distortion free effective FoV of a 2-D FT to cover the full main beam of the station in a synthesis array by second order correction was pioneered by [Bunton, private communication] who used a Gaussian convolution function with complex width parameter. This approach allows efficient imaging processing by using a 2-D FFT together with a small complex convolution kernel that provides resampling as described in subsection 3.3.2 and non-planarity correction at the same time. The so called W-projection method described by [Cornwell, 2008] uses a numerical Fourier transform for a complex beam pattern that is supposed to cor-

rect also for higher order terms, but no derivation of the required extent for the imaginary convolution kernel was provided.

In subsection 3.4.1 we extend the 2nd order analysis of Bunton and it will be shown how the phase errors that are proportional to the square of the distance from the n -axis towards the fringe tracking centre are corrected by the imaginary part of the complex width parameter of a Gaussian convolution function.

In addition it will be shown in subsection 3.4.2 how the real part of the complex width parameter limits the FoV as well as limits the extent of the required convolution kernel. The linear extent of the kernel is proportional to non-planarity and FoV.

In subsection 3.4.3 it will be shown how the very extent of the convolution kernel determines the processing cost of the convolution operation and how it drives processing efficient 2-D FTT imaging to either small FoV or to low non-planarity or to a combination of both. A reference frame with W -axis towards local Zenith of the centre of the array has low intrinsic non-planarity as determined by Earth curvature in contrast to the large extrinsic non-planarity of the conventional reference frame with W -axis towards the centre of the field of interest. The latter approach used in conventional synthesis imaging covering the full FoV of the station beam by W -projection would lead for LOFAR to excessive processing capacity for convolution far exceeding the processing capacity required for correlation.

Subsection 3.4.4 estimates an upper limit for the higher order phase residuals after convolution correction of second order terms. These phase residuals determine the actual variations in the side lobe pattern of a point source depending on its location within a 2-D FFT image. A limit on the tolerated deviations then defines an effective FoV after convolutional correction.

Subsection 3.4.5 will show that even snapshot imaging in an array based reference frame needs a too large complex convolution kernel to correct the full FoV of a 32 m LOFAR station at 50 MHz at distances further than 80 km from the array centre. Efficient processing needs an additional faceting approach, for which a Fast Faceting algorithm is presented. Such a faceting approach may even be needed for shorter distances as an efficient means to correct for direction dependent phase errors within a station beam as will occur due to e.g. ionosphere effects.

3.4.1 Quasi-convolution correction and W -projection

For an arbitrarily oriented array we can use the spherical projection approach with the W -axis towards the field of interest, which makes $n_k \sim 1$. The exponential term in expression (3.12) for G_{ijk} could then be evaluated using a series expansion for n_k according that is valid for small l_k and m_k .

By retaining the first two terms from the series expansion we get for G_{ijk} the following expression

$$G_{ijk} = G(l_k, m_k, W_{ij}) = \exp(2\pi i W_{ij}) \exp(\pi i (l_k^2 + m_k^2) W_{ij}) \quad (3.38)$$

This formula shows a fixed phase term $2\pi W_{ij}$ independent of l_k and m_k that could be brought in front of the Fourier sum provided by (3.11) and each visibility can thus be simply corrected by the fringe stopping process.

Following [Bunton, private communication] we recognize in the second factor of (3.38) a Gaussian function

$$g(r, \sigma) = \exp(-r^2 / 2\sigma^2) \quad (3.39)$$

with

$$r^2 = l_k^2 + m_k^2$$

and

$$1/2\sigma^2 = -\pi i W_{ij}$$

This suggests to obtain for C_{pq} in (3.35) also a Gaussian filter function but with a complex width parameter σ_c given by

$$1/2\sigma_c^2 = 1/2\sigma_r^2 + \pi i W_{ij} \quad (3.40)$$

The imaginary part with W_{ij} then corrects for the phase error introduced by the non-planarity of each specific U_{ij}, V_{ij} sample while the real part with σ_r provides the spatial convolution filter needed for the resampling. We need then for $C(U, V)$ the Fourier transform of a Gaussian (3.39) with complex width parameter σ_c which is also a Gaussian

$$C(U, V, \sigma_w) = (2\pi)^{-1/2} \sigma_w^{-1} \exp(-R^2 / 2\sigma_w^2) \quad (3.41)$$

with

$$R^2 = U^2 + V^2$$

and

$$\sigma_w = (2\pi\sigma_c)^{-1}$$

We use (3.41) as the convolution kernel in (3.35) to provide spatial filter and position dependent phase correction at the same time. We then find power flux density P_k on the l, m -plane according to (3.14) by including the convolution term

$$(n_k^{-1} g_k^p A(l_k, m_k) P_k) * \underline{S}_k = \sum_{ij}^N c_{ij}^m * C(U_{pq}, V_{pq}, \sigma_{ij}) \exp(2\pi i (U_{pq} l_k + V_{pq} m_k)) \quad (3.42)$$

where N is the total number of regridded samples. We need to realize however that the operation $* C(U_{pq}, V_{pq}, \sigma_{ij})$ is not a true convolution since the parameter $\sigma_{ij} = 1/2\pi\sigma_c$ is not a simple constant such as σ , but contains a term W_{ij} that depends on which baseline is actually used for v_{ij} for which the proper convolution function has to be determined. This leads not to a \underline{C}_{im} as the inverse FT of C_{pq} , but to an ampli-

tude function $A(l_k, m_k)$ that resembles \underline{C}_{lm} . Actually we need some averaging over all W_{ij} terms, which needs to be evaluated and then allows removal from (3.42) together with $n_k^{-1} g_k^p$ to obtain P_k .

In practice we could for instance start by defining \underline{C}_{lmk} as the product of two complex factors $\underline{C}_{ijk} \underline{C}_{jik}$. Each factor has a real amplitude factor, e.g. the square root of a prolate spheroidal function and a complex phase term for each telescope derived from G_{ijk}^* (3.12), but without expansion as used in (3.38). After a numerical FT, we get the correcting complex –voltage– convolution terms C_{ipq} for each antenna station. Per baseline we then need a double convolutions $C_{ipq} * C_{ij}^m * C_{jpq}^*$ where the order of the convolutions is important only when the full polarization matrices are used.

For our further analytic analysis we continue with the Gaussian function that corrects only for the quadratic terms.

3.4.2 Support of the quasi-convolution kernel

The full expression for the complex Gaussian convolution function with insertion of (3.40) in (3.41) is given by

$$C(U, V, W_{ij}) = (1 / 2\pi\sigma_r^2 + i W_{ij})^{-1/2} \exp(-2\pi^2\sigma_r^2 R^2 (1 - 2\pi i \sigma_r^2 W_{ij}) / (1 + 4\pi^2\sigma_r^4 W_{ij}^2)) \quad (3.43)$$

where W_{ij} is not an independent variable, but a parameter that depends on which stations are used to form baseline (U_{ij}, V_{ij}) .

It is important to analyse three extreme cases

i) For $W_{ij} \ll (2\pi\sigma_r^2)^{-1}$ (3.43) simplifies to

$$C(U, V, W_{ij}) = (2\pi)^{1/2} \sigma_r \exp(-2\pi^2\sigma_r^2 R^2) \quad (3.43a)$$

Which is just the Fourier transform of a Gaussian beam $C(l, m)$ with real width parameter σ_r as expected.

ii) For $W_{ij} \gg (2\pi\sigma_r^2)^{-1}$ (3.43) simplifies to

$$C(U, V, W_{ij}) = (i - 1) (2 W_{ij})^{-1/2} \exp(-R^2 / (2\sigma_r^2 W_{ij}^2)) \exp(+ \pi i R^2 / W_{ij}) \quad (3.43b)$$

iii) For $R > R_c$, using cut-off baseline $R_c = c_c \sigma_r W_{ij}$ with the scale parameter c_c , we can even ignore the whole convolution term.

The exponent with the real part containing $2\sigma_r^2 W_{ij}^2$ drives the convolution kernel rapidly down to any required low level given by $\exp(-c_c^2/2)$ for an appropriate choice of c_c . The important result of the two limiting cases i) and ii) is that the potential singularity for $W = 0$ does not exist, as is ignored in comparable analyses [Cornwell, 2008], [Humphreys, 2011].

Even more important, we can derive a maximum kernel size as follows. We define the extent of the convolution kernel $B_c = 2 \lambda R_c$ as the baseline length in meters over which the convolution needs to be extended and relate it to the station diameter D and to the non-planarity given by $H = \lambda W$ for wavelength λ . The extent of the convolution kernel for each station is then given by

$$B_c = 2 c_c \sigma_r H \quad [\text{m}] \quad (3.44)$$

A FoV determined by the station beam could use a resampling convolution that results into a Gaussian beam in the image plane that has the same width as the station beam and then requires for a typical tapering of the aperture of the phased array a standard deviation $\sigma_r = (1.2 \lambda / D)/2.36$, which then leads to

$$B_c = c_c \lambda H / D \quad [\text{m}] \quad (3.44a)$$

For $c_c = 5$ we get a cut-off for the convolution kernel at $\sim 10^{-6}$, which limits the error caused by ignoring contributions that fall outside the kernel extent [Brouw, 1974], and we give a first order estimate for the kernel extent K_c in pixels using a typical grid spacing $S_g = D / 2.8$ as

$$K_c = B_c / S_g = 14 (\lambda / D) H / D \quad (3.45)$$

In practice a prolate spheroidal function [Humphreys, 2011] is used instead of a Gaussian that has steeper decay which results into a smaller effective c_c . This potential advantage could however be offset by choosing a grid spacing smaller than $D/2.8$, so we use the number 14 in the equation as representative for our further analysis.

Using σ_r smaller than half the station beam width creates a facet beam that is smaller than the station beam and needs a larger convolution kernel. Since the FFT field could be reduced as well, the kernel keeps the same size in grid units.

3.4.3 Comparison with W-projection analysis and discussion

We compare our analysis with the proposed W-projection approach [Cornwell, 2008] where the extent of the quasi convolution kernel is not explicitly derived. In a recent paper convolutional resampling is discussed [Humphreys, 2011] and an

approximate kernel extent is argued that is consistent with our results. We summarize our approach and point out the differences with Cornwell and with Humphreys.

Section 3.3.1 introduced a resampling convolution kernel C_{pq} that relates the projection of measured visibilities c_{ij}^m defined by (3.11) on the U,V-plane to regridded visibilities c_{pq}^r defined by (3.34). This convolution kernel is the 2-D Fourier transform of C_{lm} in (3.37) which equation defines the relation between the sky image I^s and the image I_{lm}^r obtained after 2-D Fourier inversion of the regridded visibilities c_{pq}^r . If the convolution kernel C_{pq} would be chosen such that C_{lm} includes an imaginary term that is the inverse of G_{ijk} introduced in (3.12) then (3.37) can be simplified to (3.37a).

We simplified our analysis by introducing a Gaussian convolution kernel that uses a complex width parameter [Bunton, private communication] which has a Fourier inverse that is also a Gaussian and we assumed that it has a complex width parameter that is just inversely proportional to the complex width of the convolution kernel. By proper choice of the complex width parameter in C_{lm} for each baseline sample that needs to be convolved we get an imaginary part that just cancels the terms with $l^2 + m^2$ in an expansion of the exponent in G_{ijk} .

Cornwell et al. simply used an imaginary width parameter that has a singularity for $W = 0$ in the assumed Gaussian convolution function C_{pq} which indicates that their explanation is inconsistent and raises doubt on the approach. Humphreys et al. still ignore this singularity and focus on the real part of C_{pq} that is *sinc* shaped and has a slow decay. Nevertheless their proposed cut-off for a limited kernel extent agrees reasonably well with our derivation and imaging results show indeed little distortion as predicted by our analysis and are confirmed in simulations [Labropoulos, private communication].

Cornwell et al. attempt to generalize our limited approach by deriving a gridding kernel as the Fourier inverse of the product of the complex conjugate of G_{ijk} with a real prolate spheroidal function. There is however no published proof that this approach indeed supports more terms of an expansion of $(1 - l^2 - m^2)^{1/2}$ in the exponent. In our analysis we consider only correction for the quadratic terms and our result is that higher order terms still produce phase errors on certain baselines and give distortion of sources. The advantage of our analysis is that we can give an upper bound for these distortion that define the FoV of 2-D Fourier image of a non-planar array as will be discussed in subsection 3.4.5. These remaining distortions are well quantified by an upper bound and can be compared with other distortions, for instance from long integration of a snapshot image depending on the reference coordinate system and on field rotation as will be discussed in section 3.5, and by calibration as will be discussed in chapter 4.

3.4.4 Convolution processing determined by choice of U,V-reference plane

Synthesis imaging using a single 2-D FFT needs a coordinate system with the W-axis towards the source field but gets a large extrinsic non-planarity defined by the projection of the interferometer baselines on the W-axis. A reasonable estimate for the maximum non-planarity for a tracking period larger than 6 hours is $H_{\max} \sim B_{\max}/2$ and we can insert this in (3.45) and find for the maximum kernel diameter that corrects for extrinsic non-planarity

$$K_{\max}^{\text{Ex}} = 7 (\lambda / D) (B_{\max} / D) \quad (3.46)$$

This equation shows that K_{\max}^{Ex} is proportional to the diameter of the FoV and to the number of resolution elements over that diameter. Another arrangement of the variables is given by

$$K_{\max}^{\text{Ex}} = 7 (\lambda / D)^2 (B_{\max} / \lambda) \quad (3.46a)$$

This form shows that the kernel diameter is proportional to the FoV in sr and to the baseline expressed in wavelength and is not surprisingly equal to formula (3.10) for the number of planes in the 3-D FFT approach except for the factor 7.

We take now as example the maximum beam width situation for LOFAR obtained at 50 MHz with the small LBA configuration that has a diameter of 32 m. For $B_{\max} \sim 6$ km we find $K_{\max}^{\text{Ex}} \sim 250$.

The processing capacity N_k in Complex Multiply Add (CMA) operations that is required to convolve a single visibility sample to all pixels on the visibility grid within the extent of the kernel is then given by $N_k = K_c^2$. In our widest beam case example for LOFAR that corrects for extrinsic non-planarity we then need $\sim 0.6 \cdot 10^6$ CMA for each visibility sample of the longest baseline.

The number N_k could now be compared with the number of CMA operations required for the correlation of such a single visibility sample in an FX correlation system, which is given by

$$N_c = \tau \delta\nu \quad (3.47)$$

For baselines of 6 km we need according to table 3.2 an integration time $\tau = 12$ s and according to (3.33) a bandwidth $\delta\nu = 42$ kHz and we find $N_c \sim 0.5 \cdot 10^6$ CMA.

From the viewpoint of system optimisation it is reasonable to assign equal budget [Bregman, 2004a] to correlation and image forming platforms. Since LOFAR uses comparable types of High Performance Computing facilities for both platforms, we can do a simple cost comparison based on CMA count. This implies that the pro-

cessing platform that should do the convolution for each visibility sample is of comparable cost to the platform for correlation, just to keep up with the data stream of visibility samples. Recent implementation on a GPU based HPC platform shows that this might be a cost effective solution for kernels up to 250^2 [private communication, Labropoulos] but cannot handle the longer baselines for LOFAR.

Alternatively, we could correct only for intrinsic non-planarity caused by Earth curvature. We then need a coordinate system with W -axis toward local Zenith of the centre of the array. Insertion of (3.26) in (3.45) makes K_c dependent on the maximum distance L_{\max} of a station from the centre of the array and we find for the maximum kernel diameter that corrects for intrinsic non-planarity

$$K_{\max}^{\text{in}} = 7 (\lambda / R_E) (L_{\max} / D)^2 \quad (3.48)$$

The kernel diameter now scales with $(L_{\max} / D)^2$ instead of (B_{\max} / D) as in (3.46) but has a larger reduction factor (λ / R_E) instead of (λ / D) .

An important observation is that for a given array configuration the kernel diameter scales just as in (3.46) with wavelength. Even more important is that the kernel diameter scales for intrinsic and for extrinsic non-planarity with D^{-2} and important processing savings could therefore be realized if the FoV could be reduced.

Although the convolution processing is now proportional to the 4th power of the distance L from the centre of the array, a large reduction factor has entered the equation, which reduces the processing requirements dramatically. Our wide beam situation for LOFAR operating at 50 MHz using LBA stations in the small 32 m configuration needs for stations out to 80 km from the centre of the array $K_{\max}^{\text{in}} \sim 41$ requiring $N_k \sim 1700$ CMA. We need 2.8 s integration time and channels of 3 kHz requiring 8400 CMA per visibility, making convolution processing feasible in principle.

However, baselines to 68 m stations at 600 km from the centre of the array need a large convolution kernel of 512^2 for each sample. Even more serious is that each correlated sample of $\tau = 0.13$ s integration time and $\delta\nu = 0.5$ kHz bandwidth needs to be convolved for imaging of a full station FoV at the full resolution of the 1200 km baselines. A single visibility sample then requires only 65 CMA for correlation but convolution would require $0.26 \cdot 10^6$ CMA, which is not affordable as discussed above.

The following preliminary conclusions can now be drawn:

- Full beam FoV imaging with a 2-D FFT using convolution correction for extrinsic non-planarity appearing in the conventional coordinate system with W -axis towards the source, requires for a maximum baseline of ~ 6 km

- more processing power per visibility sample for convolution than for correlation. This convolution power increases proportional to the square of the baseline and cannot be afforded in practice.
- Full beam FoV imaging using convolution correction for stations out to ~ 80 km from the centre of the array can be computationally afforded only if intrinsic non-planarity needs to be corrected, as for instance by Earth curvature.
 - This could in principle be realized by choosing a Cartesian coordinate system with W-axis toward local Zenith of the centre of the array for 2-D Fourier snapshot imaging in the U,V-plane, and will be further analysed in section 3.5.
 - For stations at distances larger than ~ 80 km from the core we need however to limit the extent of the convolution kernel by reducing σ_r in (3.44), which reduces the FoV.
 - The FoV is then no longer determined by the station beam but by the narrower “gridding” beam that has a width that corresponds to a larger “virtual station” diameter.
 - This allows longer integration time for the baseline samples and for continuum imaging also larger channel bandwidth, which together decrease the number of input samples for an FFT snapshot image within the FoV of the station beam.
 - A large FoV is then obtained by processing of a large number of small fields within the station beam and will be further discussed in section 3.4.6.

3.4.5 Field-of-view of a 2-D Fourier image after complex quasi-convolution

In section 3.1.8 we derived expression (3.25) for radius $\Delta\theta_r$ of the FoV as determined by the maximum W value in a 2-D FT image of a synthesis array. We derived (3.25a) for radius $\Delta\theta_0$ of the FoV of a quasi-planar array and both formulae stem from second term and third term between parentheses in (3.23) respectively. As the 2nd order terms can be corrected the remaining phase error after correction is given by

$$\delta\varphi_c = \pi (l_s l_0 + m_s m_0)^2 (H / \lambda) n_0^{-3} \quad (3.49)$$

where l_0 and m_0 are the direction cosines of the centre of the field with a source at distance (l_s, m_s) from that centre. There are two situations of practical importance, (i) $|l_0| \sim |m_0| \sim |l_s| \sim |m_s| \sim \Delta\theta_{rc}$ and (ii) $|l_0| \sim |l_s| \sim \Delta\theta_{oc}$ while $|m_s| \sim \Delta\theta_{oc} n_0$. The radius of the correctable FoV $\Delta\theta_c$ is either given by distance $\Delta\theta_{rc}$ from the W-axis or by distance $\Delta\theta_{oc}$ from a nominal position at $l_0 \sim 0$ by appropriate rotation of the U,V-coordinates in the reference plane of a quasi-planar array.

i) The first case is also applicable to the situation with W-axis to the centre of the field with $n_0 = 1$, since the neglected 4th order terms that need to be taken into account are approximately covered by

$$\delta\varphi_c \sim \pi \Delta\theta_{rc}^4 (H / \lambda) \quad (3.49a)$$

If we again use $\delta\varphi_c = \pi^{-1}$ we get after 2nd order correction a FoV with radius $\Delta\theta_{rc}$ >> $\Delta\theta_r$ given by

$$\Delta\theta_{rc} \sim \pi^{-1/2} (\lambda / H)^{1/4} \quad (3.50)$$

ii) Our second case with large offset m_0 from the W-axis gives additional second order terms in l_s and m_s that are corrected as well but there remains a cross term giving

$$\delta\varphi_c \sim 2\pi \Delta\theta_{0c}^3 (H / \lambda) m_0 n_0^{-2} \quad (3.49b)$$

Using the standard assumption $\delta\varphi_c = \pi^{-1}$ we get

$$\Delta\theta_{0c} \sim (2\pi)^{-1/3} (\lambda / H)^{1/3} n_0^{2/3} m_0^{-1/3} \quad (3.50a)$$

The FoV size as determined by the not-corrected higher order terms actually needs a certain kernel size of a complex Gaussian to corrects for the 2nd order terms in l_s and m_s that would otherwise reduce the FoV. The maximum FoV of a beam defines the minimum diameter of the aperture that needs to be corrected and is therefore in case (i), applicable to extrinsic non-planarity $H = B_{\max}/2$, given by

$$D_{\min}^E \sim 0.6 \lambda / \Delta\theta_{rc} \sim 0.9 \lambda (B_{\max} / \lambda)^{1/4} \quad (3.50b)$$

For case (ii), applicable to intrinsic non-planarity, the minimum aperture diameter is given by

$$D_{\min}^I \sim 0.6 \lambda / \Delta\theta_{0c} \sim 1.8 \lambda (H / \lambda)^{1/3} n_0^{-2/3} m_0^{1/3} \quad (3.50c)$$

If $D_{\min} > D$ we are in a situation that higher order terms dominate the tolerated phase errors in imaging although 2nd order terms are corrected by Gaussian convolution, which means that faceting is required to cover the full station beam. Using smaller facet beams with aperture diameter $D_f > D_{\min}$ the higher order phase terms will be reduced progressively in a 2-D Fourier image with a non-planar array.

If we take the same examples from subsection 3.1.9 using a FoV radius of 0.11 at a wavelength of 6 m, we find after correction for second order terms for the intrinsic case (ii) with $n_0 = m_0 = 0.7$ a value $H_{\max} \sim 502$ m that allows $L_{\max} = 80$ km instead of 30 km. The extrinsic case (i) now allows in principle a maximum baseline of 7 km

instead of 72 m, which is not enough for LOFAR but, but even more seriously, the required convolution kernel size is computationally not affordable as discussed in subsection 3.4.4.

Although a snapshot imaging approach without faceting is a potential option for the Dutch LOFAR configuration, it is now clear that some form of faceting is needed anyhow to reach baselines up to 1200 km. The minimum number of facets for the beam of a station with aperture diameter D is then given by $N_{\min} = (D_{\min} / D)^2$.

3.4.6 Fast Facet imaging

In section 3.4.4 it was shown that convolution correction for extrinsic non-planarity needs for baselines of 6 km already excessive processing resources. But even correction for just intrinsic non-planarity to obtain the full FoV of the station beam for international LOFAR stations at 600 km distance from the array centre would also require an excessively large convolution kernel. More serious is that the convolution should be applied to a huge data stream of visibility samples with very short integration time and very narrow bandwidth.

For our situation, we choose snapshot imaging in a coordinate system where only intrinsic non-planarity needs to be corrected. Evaluation of (3.48a) shows that the factor $(\lambda/D)^2$ is the one that effectively determines the required convolution processing capacity, since the other parameters leave no alternative choice. We could therefore choose to image a smaller facet beam that corresponds to a larger virtual station diameter rather than the full FoV provided by the station main beam.

We analyse the scaling of a synthesis array that has outermost stations at 600 km from the centre instead of 60 km. Our facet field needs a 10 times smaller angular diameter but has a 10 times higher spatial resolution, so the number of resolution elements in the facet image is the same as in the full FoV of the small array. We could therefore use the same bandwidth and integration time of order 4 kHz and 1 s respectively to get an acceptable degradation on the longest baselines for sources at half power of the smaller facet beam (that will however be corrected for).

Although the convolution kernel would have a 10 times wider extent D/λ in the U,V-domain we still have the same number of pixels in the kernel, since our U,V-grid has become coarser as well. Full FoV imaging of the station beam then needs imaging of 100 facets, which just increases the total processing capacity for convolution and Fourier transformation with the same factor. This shows that in a faceting approach the total processing capacity for convolution and Fourier imaging is just proportional to the total FoV expressed in resolution elements and is independent of actual FoV or resolution. However, the total number of required facets depends on the choice of the coordinate system used for the 2-D Fourier transforms.

The faceting approach needs a visibility dataset that tracks the centre of each facet, which suggests an increase in required correlation processing capacity. Although current multiple fields processing of VLBI observations indeed uses multiple correlation passes we do not need full reprocessing but only dedicated fringe stopping for each facet, which could be realized after correlation. Polyhedron imaging is an actual implementation of faceting that reprocesses a given visibility dataset a number of times leading to large processing costs. However the chosen implementation is not optimum and lacks convolutional correction for non-planarity, but can be improved by including a small but complex convolution kernel and using alternative facet forming as will be explained as follows.

A relevant question is how could we provide up to ~ 100 visibility datasets that each track their own facet? The LOFAR correlation design assumed a maximum data output rate based on a single dataset with 1 second integration time and 1 kHz channel bandwidth, although 4 kHz would be matched in terms of bandwidth and integration time smearing at 50 MHz. If we make for instance facet datasets with samples that have 1 s integration time but 4 kHz bandwidth, we could easily stream 4 datasets in parallel through the output ports. At an observing frequency of 150 MHz we need only 12 kHz channels, which even allow 12 data streams that cover the same total bandwidth.

We need to realize that smaller facets do not require the bandwidth and integration time that limits smearing for fields that cover the main lobe of the station beam. Instead, a facet with half the diameter of the station beam allows double bandwidth and double integration time. So, 4 facets that fill the station beam produce 4 data streams each filling $1/4^{\text{th}}$ of the nominal data rate. This process can be repeated and shows that faceting does not increase the total data rate.

Instead of shifting each sample in one go to all facet datasets, an FFT like butterfly approach should be used, where intermediate facet datasets are created and each sample get shifted in a number of steps.

A processing-efficient Fast Faceting scheme uses the following steps:

- We start with a basic visibility dataset that contains a time series of spectra, i.e. a 2-D array.
- The array consists of tiles each with 4 adjacent elements that have odd and even numbered elements along time and frequency axis as depicted in the left side of figure 3.4. The odd and an even time stamps t_o and t_e mark the centres of two subsequent integration intervals while ν_o and ν_e mark the centres of two adjacent spectral channels of each baseline
- In our example we have samples with integration time $\delta\tau = 0.06$ s and bandwidth $\delta\nu = 0.24$ kHz that describe the full beam area of $\sim 13^\circ$ FWHM

- centred at (l_0, m_0) of a station of 32 m diameter operating at 50 MHz for a baseline of 1200 km.
- We define a new visibility dataset with tiles where each of the four elements represent a sample fringe stopped for $(l_0 - \Delta l_0, m_0 + \Delta m_0)$, $(l_0 + \Delta l_0, m_0 + \Delta m_0)$, $(l_0 - \Delta l_0, m_0 - \Delta m_0)$, and $(l_0 + \Delta l_0, m_0 - \Delta m_0)$ respectively as depicted in the right side of figure 3.4.
 - Each visibility sample in the 4 element tile at the left side is shifted 4 times and summed with the three other samples in its tile and stored into an element of the new tile at the right side.

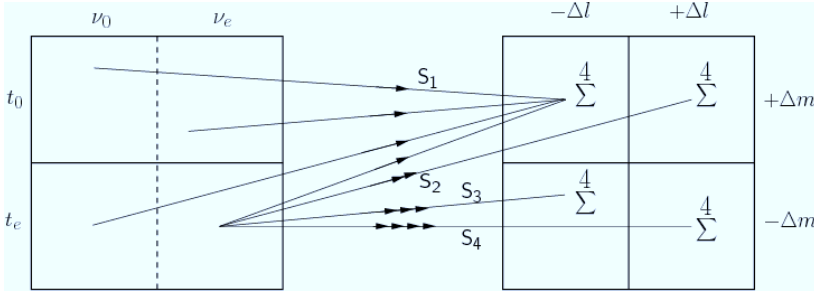


Figure 3.4. Butterfly kernel of the Fast Faceting algorithm showing constant data volume.

- The fringe position of each new sample is at the centre of a quadrant of the original field and has the proper phase for $(t_0 + t_e)/2$ and $(\nu_0 + \nu_e)/2$.
- All tiles in the left data set are processed successively. And we end with four separate facet datasets.
- This process is repeated where four adjacent samples of each facet dataset are distributed over 4 smaller facet datasets and we get a visibility sample with four times the initial integration time and four times the initial bandwidth for each of the 16 smaller facets.
- The number of visibilities per facet dataset decreases but the number of facet datasets increases, while the total amount of data remains the same.
- In every step we half the effective diameter of each facet FoV, we quadruple the number of facets and we require 4 CMA for each original correlation sample.
- After n steps we reduced the diameter of the FoV per facet to 2^{-n} times the diameter of the station beam and generated 4^n datasets.
- The total number of CMA per original correlation sample is $\tau \delta \nu = 14$ and we needed an additional $4n$ CMA for each original data sample.
- After 4 steps we used only 16 CMA per original data sample but have got 256 datasets with samples of 1 s integration time and 4 kHz bandwidth that require a small convolution kernel.

- We do not need large additional storage capacity, since processing is per 4 elements of a tile and the results for the new 4 element tile can be stored at the locations of the old tile.
- For less than 16 facet datasets a conventional approach is more efficient than the butterfly approach.

The data rate of the stream of 0.24 kHz channels of 0.06 s integration is much greater than the nominal data output rate of the cross-correlation processing based on samples of 1 kHz and 1 s. The fast faceting could produce 256 separate datasets of 1 s and 4 kHz but the available output data rate supports only 4 facet datasets that cover only 1.6 % of the beam of a compact LBA station but 6 % of the beam of a larger European station. We can select only 4 out of the 256 available datasets for further processing, but we need at least 5 fields with strong calibrators that allow ionospheric phase calibration. Next to the astronomical relevant datasets we need additional ones that contain sources that need to be subtracted from the astronomical fields. The number of facets should therefore be increased by continuing the fast faceting process 3 more steps providing a total of 16384 facet datasets of which 256 facets centred on relevant objects could be selected for further processing. The facets become smaller and have visibility samples of 8 s and 32 kHz for conveniently sized facet images of 4096^2 pixels. From each facet only the central 2048^2 pixels are retained, which cover an area of $0.1^\circ \times 0.1^\circ$ that has no aliased artefacts and only little noise enhancement at its edges. The FWHM of the resolution beam is sampled by 3 pixels while 8 s of sky rotation corresponds to a shift at the edge of the field of most 0.6 pixel. Since 8 s integration time also means resampling of the U,V-tracks a spoke like side lobe pattern will arise at distances $> 0.12^\circ$ from point sources, but falls outside the retained facet image.

Before the 256 selected datasets are actually transferred from the correlation platform to the storage platform they could be compressed a factor 4 by averaging 2 time and 2 frequency channels, which increases the decorrelation from 1.7 % to 7 % for sources at half power of the facet beam on baselines with the largest non-planarity. This allows even 1024 compressed facet datasets with samples of 16 s and 64 kHz to be transferred and allows covering up to 24 % of the main beam of the European stations. The spoke like side lobe pattern will now start at distances $> 0.06^\circ$ from point sources. At 150 MHz even 72 % could be covered since the total number of integrated spectral channels in each facet image is a factor 3 lower.

We showed that the long baselines between the European stations that need high temporal and spectral resolution can in principle be handled by the existing correlation platform by forming a large number of facet datasets with lower resolutions of which a fraction is actually transferred to another platform. This fraction could however cover 24 % of the beam area defined by the FWHM of the European stations at 50 MHz and even 72 % at 150 MHz. The 7 % sensitivity loss at the longest baselines for sources at the edges of the final facet images is small compared to the taper value that is normally used in imaging with these longest baselines.

Of course this increase of integration time and bandwidth together only works for continuum observations. For line observations the band width should not exceed a level determined by the application. However, faceting with larger integration times is still possible, but increases the data volume since the spectral axis is not reduced in number of samples. This approach could be used in an imaging package, but is not attractive for reducing data rates from the correlation platform.

The actual facet size that is needed depends on the choice of the synthesis imaging approach. A snapshot approach that needs only correction for intrinsic non-planarity allows a limited number of large facets to get a sufficiently small convolution kernel, but requires a number of snapshots. A synthesis approach with extrinsic non-planarity needs more facets, and a comparison taking into account the processing balance between convolution operations and Fourier operations will be made in section 3.7.

In subsection 3.3.2 we identified the importance of bandwidth and integration time smearing especially for suppression of objects outside the FoV. Simple block integration leads to a highly varying attenuation with slow decay as function of distance. For instance, integration over a set of contiguous samples and giving them a triangular weight over the interval, would transform the sinc attenuation function for uniform distribution into a sinc^2 one that is much more effective in suppressing distant objects. Such a scheme needs a modified butterfly approach, and could even need interleaved samples to preserve sensitivity at the expense of increased data rate and increased processing for imaging.

It seems logical to combine this fast facet dataset generation with flagging of bad data samples at the lowest level. However, deleting samples changes the average time and frequency of a sample. This is most easily prevented by deleting an additional sample with a symmetric position relative to the expected time-frequency average. Deleting t_o, v_o is compensated by also deleting t_e, v_e .

Further discussion on processing aspects will be given in section 3.7.

3.4.7 Summary, Conclusions, and Results

We summarize the results and conclusions of the different subsections:

- In subsection 3.4.1 we have shown how the phase distortions in the visibilities of a non-planar array can be corrected by a convolution operation such that a 2-D Fourier transform produces a distortion free image.
 - An analysis is given for a Gaussian convolution with a complex width parameter that only corrects phase errors that are propor-

- tional to non-planarity and to the square of the distance of a source from the W-axis.
- In subsection 3.4.2 the required extent of the convolution kernel has been estimated, which is verified against available results.
 - In subsection 3.4.3 our results are compared with published results of the W-projection method that used an inconsistent derivation of the required kernel size.
- In subsection 3.4.4 the processing requirements for a convolution correction of a visibility sample are compared with the processing required by cross-correlation of that sample.
 - Conventional 2-D synthesis imaging has large extrinsic non-planarity that is on average equal to half the maximum baseline and requires for full station beam correction a convolution that is too large to be feasible for LOFAR.
 - Snapshot imaging with a quasi-planar 2-D array has much smaller intrinsic non-planarity as determined by Earth curvature and requires a much smaller convolution kernel, which makes a single FoV image feasible at 50 MHz for the full beam of 32 m stations out to 80 km from the centre of the array.
 - Unfortunately, including European stations makes the convolution kernels also in this case too large to be feasible and requires a number of small facet images to cover sufficient distortion free FoV.
 - Section 3.4.5 estimated the magnitude of the higher order phase errors after correction of second order terms by complex Gaussian convolution and derived estimates for the distortion free FoV needed to define the number of required facets.
 - Conventional 2-D synthesis imaging has a FoV diameter that scales with extrinsic non-planarity to the power 1/4 which makes faceting feasible and will be discussed in section 3.7.
 - Snapshot imaging with a quasi-planar 2-D array has a FoV diameter that scales with intrinsic non-planarity to the power 1/3.
 - Although no facets are required for arrays with radius smaller than 80 km, there is an additional third order term that limits the maximum tracking time of a synthesised snapshot image as will be discussed in section 3.5
 - An minimum aperture diameter D_{\min} for the facet beam is derived that defines the size of the required complex convolution kernel for 2nd order correction.
 - Smaller facet beams with aperture diameter $D_{\text{fac}} > D_{\min}$ give progressively reduced higher order phase errors in a 2-D Fourier image of a non-planar array.

- A Fast Faceting algorithm has been presented in section 3.4.6 that easily generates 16,384 facets per beam of a 32 m station at 50 MHz to handle LBA stations up to 600 km from the centre of the array and baselines up to 1200 km.
 - The processing power required for fringe shifting then equals the processing power for correlation alone.
 - About 1024 facets with samples that are compressed by a factor 2 to 16 s integration time and 64 kHz bandwidth can be transferred as limited by current available output data rate of the correlation platform of LOFAR (that can however be extended).
 - Each facet produces finally a facet image with 2048^2 pixels covering $0.1^\circ \times 0.1^\circ$ on the sky and all facets together fill $1/4^{\text{th}}$ of a European station beam.
 - The compression increases the decorrelation of the visibilities on the longest baselines for sources at the edge of the facet from $\sim 1.7\%$ to $\sim 7\%$, which in fact means an additional taper and broadening of these sources.
 - Conventional facet fringe shifting could have provided only 16 facets with the same processing power.
 - A configuration containing only the Dutch stations could for instance benefit from subdivision of the full FoV of a LOFAR station beam into 64 facets with visibility samples of 32 kHz at 8 s intervals and could then support 2-D synthesis with corrected extrinsic non-planarity.
 - This facet approach simplifies applying the appropriate direction dependent phase corrections for e.g. ionosphere disturbance just per facet image as will be discussed in chapter 4.
- In each facet that is Fourier imaged we suffer from side lobes of strong sources in surrounding facets. These side lobes are mainly caused by limited interferometric sampling as discussed in subsections 3.3.1 and 3.3.2, and can only be removed by subtracting each disturbing source from the visibility dataset of each facet that needs to be imaged.
 - We need not only the facet datasets with objects of interest but also the datasets with sources that need to be used in the multi-direction self-calibration.
 - All sources in the station beam which are so strong that they need to be subtracted correctly from all facet datasets that need to be imaged, need to be available either by a sky model or by a facet dataset.
 - Sources outside the facet image could in principle be attenuated using advanced bandwidth and integration time smearing using some weighing when samples are integrated to larger bandwidth and longer integration time.

- Enhanced bandwidth and time smearing effects then complicate the subtraction process, which is the penalty for the large decrease in captured data volume.
- Proper modelling of the calibration sources and the disturbing sources requires only a very small facet that is reasonably centred on the source, and uses a correspondingly small facet visibility dataset. Such datasets could for instance be processed with a conventional FT instead of an FFT.

The main conclusions and results for FoV extension of 2-D Fourier imaging with non-coplanar baselines are

- Conventional methods like W-projection and polyhedron imaging need too much processing power to be of practical use for LOFAR.
- We derived a relation for the size of a complex quasi-convolution kernel that allows pre-processing of non-coplanar baseline visibilities such that 2nd order terms in 2-D Fourier imaging are fully corrected.
- Using a minimum size for such a kernel, we find a maximum FoV in 2-D Fourier imaging, limited by higher order terms, which requires for LOFAR a large number of small facet images to fill a complete station beam.
- We developed a so-called Fast Faceting method that minimizes pre-processing for such a large set of small Fourier images.
- For snapshot imaging where the non-planarity is only determined by Earth curvature the maximum FoV of the Dutch LOFAR configuration can be handled by a single large Fourier image.

Limitations for a long synthesis formed by snapshot images will be the subject of the next section.

3.5 Snapshot synthesis in an array based coordinate system

In this section a procedure is considered that increases the tracking time for snapshot images. This procedure requires projection of the baselines on the chosen reference plane from the direction of the image centre, which provides first order phase corrections during the tracking interval. Complex convolution could then correct for second order effects and the third order effects are controlled by limiting the snapshot integration time and facet diameter. It will be shown that a series of synthesized snapshot images show parallactic rotation of their image coordinate

systems but require only a small correction for image rotation during their synthesis interval.

The snapshot synthesis approach allows us to describe the effects of intrinsic non-planarity caused by Earth curvature in the FoV of a 2-D Fourier image of a synthesis array with large extent and beam size such as LOFAR. Especially the effects of foreshortening deformation on the beam shape and of changing polarization characteristics due to tracking and parallactic rotation can most easily be described in the Earth bound hemispheric image domain. This domain includes the local horizon which for the core of the array is also the location where most Earth bound RFI will appear. A full hemispheric imaging range enables analysis of the effects of all sources and their distortions by the nominal point spread function (psf). Successful demonstrations of this approach are the all-sky images formed from a set of snapshot images made from the correlations between the elements of a single LOFAR LBA station [Wijnholds, 2004, 2005, 2010].

Although snapshot refers to instantaneous 2-D imaging, we will show that the 2-D Fourier image in the plane of a quasi-planar array could be extended to a short synthesis image of typically 10 min duration called a synthesized snapshot image where a sky field is tracked for a while by applying a set of simple corrections. A full synthesis of ~6 h at 50 MHz using compact LBA stations of 32 m diameter out to 80 km from the centre of the LOFAR array would need ~ 40 synthesized snapshot images, which is much less than ~ 400 Fourier planes needed by the 3-D imaging approach. The important issue is that only intrinsic non-planarity caused by Earth curvature needs to be corrected, which requires only a small but complex convolution kernel.

In subsection 3.1.6 we have shown that imaging with a non-planar array can still be described by a 2-D Fourier transform, but after a 3-D fringe shift image distortions appear for sources that are not near the fringe stopping centre and we will extend that analysis in subsection 3.5.1.

Subsection 3.5.2 analyses the impact of non-planarity on a so called snapshot image that is obtained by a 2-D Fourier transform from a set of visibilities that have been integrated over a short enough time interval that sky rotation effects can be ignored.

Subsection 3.5.3 analyses the impact of extending the integration time to obtain a synthesized snapshot and subsection 3.5.4 focuses on the rotation aspects when a sky source is tracked.

Finally subsection 3.5.5 concentrates on effects of combining synthesized snapshots that have different image scales in a full synthesis image, and we end with a summary.

3.5.1 2-D Snapshot imaging with a non-planar array

In section 3.1.6 we derived expression (3.21) for the phase of a point source as observed with an almost planar array in a coordinate system with W -axis perpendicular to the plane of the array. We will show how we could define a reference coordinate system for snapshot imaging that allows further simplification of the phase error term for sources at a distance from the fringe stop position (l_0, m_0, n_0) . We rewrite (3.21) and define corrected U_0, V_0 coordinates $U_0 = U - l_0 W_0$ and $V_0 = V - m_0 W_0$ that should be used by the 2-D Fourier transform where $W_0 = W / n_0$ and $n_0^2 = 1 - l_0^2 - m_0^2$. We then get

$$\begin{aligned} \varphi / 2\pi = & l_0 U + m_0 V + n_0 W \\ & + l_s U_0 + m_s V_0 \\ & - (l_s^2 + m_s^2) W_0 / 2 - (l_s l_0 / n_0 + m_s m_0 / n_0)^2 W_0 / 2 \end{aligned} \quad (3.51)$$

The first line shows the required fringe shift correction that places the centre of our 2-D FT image with l_s, m_s -coordinates on (l_0, m_0, n_0) defined in our U, V, W -coordinate system with W -axis towards Zenith, i.e. perpendicular to the assumed reference plane of the array.

The U_0, V_0 -coordinates in the second line of (3.51) are corrected for the tilted baselines with a projection correction for W from direction (l_0, m_0) and are in fact the projection of the baseline on the reference plane from the direction towards the centre of the field as depicted in figure 3.3.

The last row shows the familiar quadratic terms in l_s and m_s , but now with an enhanced W_0 and an additional term, which also has quadratic terms in l_s and m_s that can easily be combined with the first terms of the third line, but also has a cross term giving

$$\delta\varphi_c / 2\pi = l_s m_s (l_0 / n_0) (m_0 / n_0) W_0 \quad (3.52)$$

The result of the fringe shift operation is an effective boost of W by the elevation factor n_0^{-1} to W_0 for fields centred at large zenith angle θ_0 since $n_0 = \cos(\theta_0)$. The elevation factor enhances not only the quadratic terms in l_s and m_s but especially the additional cross error term (3.52) that dominates at low elevation. This will have limited consequence for the LOFAR synthesis array with phased array stations that will in practice not observe below 15° elevation because of the very much reduced sensitivity and increased ionosphere disturbances. Although n_0 is defined for the reference plane of the array, local horizon and elevation at each station is different.

For 2-D Fourier imaging in the plane of the array the source appears distorted according to (3.51) due to phase errors for a non-planar array, which are in addition to distortion by amplitude variation as result of bandwidth and integration time smear-

ing. With a fringe shift to a nominal position near the RFI source, the imaging phase errors are reduced and the source could be properly imaged, although still distorted due to amplitude effects by time and bandwidth smearing. For that situation it would be better if the decorrelation was not a sinc function with strong periodic attenuation but for instance a sinc^2 function by using an appropriate weighing scheme for the integration over time and bandwidth samples. Such a weighing scheme would be a minor complication for the fast faceting algorithm described in section 3.4.6.

An important observation is that the cross error term (3.52) is zero for either $l_0 = 0$ or $m_0 = 0$. The phase defined by (3.51) is invariant for 3-D rotation of the coordinate system [Sault, 1996]. Therefore, we could define a coordinate system U', V', W' that is rotated along the W -axis such that $W' = W$ and requiring that the U' -axis is pointed such that $l'_0 = 0$ and so $n'_0 = (1 - m'^2_0)^{1/2}$, while $n'_0 = n_0$ and $W'_0 = W_0$, which simplifies (3.51) with $V'_0 = (V' - m'_0 W'_0)$ and $U'_0 = U'$ to

$$\begin{aligned} \varphi / 2\pi = & + m'_0 V' + n'_0 W' \\ & + l'_s U'_0 + m'_s V'_0 \\ & - (l'^2_s + m'^2_s) W'_0 / 2 - (m'_s m'_0 / n'_0)^2 W'_0 / 2 \end{aligned} \quad (3.53)$$

The first line of (3.53) shows the total required fringe stop correction, of which a part is already done before and after correlation processing and the remainder by the imaging process. The second line defines the 2-D FT on the U'_0, V'_0 -coordinates and the third line shows only quadratic terms in l'_s and m'_s and no longer a cross term, while higher order terms are ignored. This particular choice of the coordinate system with U' -axis towards the field centre then allows a 2-D FT to obtain an image in l'_s, m'_s -coordinates.

So far we have shown that by choosing an appropriate coordinate system for a quasi-planar array with $l'_0 = 0$, we have no longer image distortions in a 2-D Fourier image by phase errors linear in l'_s and m'_s . The effective FoV could be extended using the quasi-convolution correction for terms in l'^2_s and m'^2_s as described in section 3.4. This means in effect that the FoV is limited by the higher order terms, which is the subject of the next section.

3.5.2 Sky tracking with a shifting correction for the 2-D Fourier image

The previous subsection discussed the situation that every instantaneous snapshot image made with a tracking array observes the field at different elevation defined by n_0 and in our special case by m_0 . This means that the m_s coordinate in different snapshot images corresponds to different positions in the l', m' -plane, suggesting that we need to change the reference coordinate system continuously. On the other hand it would be attractive if an array-bound coordinate system could stay fixed for some time, such that a synthesis image over a short period of time could be made with l'_s and m'_s coordinates relative to a reference l'_0, m'_0 that is constant. We there-

fore investigate the type of phase error terms that would result if we track a sky field in a coordinate system of an array that is defined for the centre of a short interval. We choose an array bound coordinate system U', V', W' that is rotated along the W -axis (towards Zenith) such that the field centre of interest is located in the U, W -plane for the middle of the tracking interval. Then $l'_0 = 0$ with $m'_0 = \sin(\theta_c)$ and $n'_0 = \cos(\theta_c)$ with zenith angle θ_c . For a moving sky we have to replace in the first line of (3.51) l_0 by $\delta l'_0$ and m_0 by $m'_0 + \delta m'_0$ but n_0 by $n''_0 = (1 - \delta l'^2_0 - (m'_0 + \delta m'_0)^2)^{1/2}$ requiring adapted fringe tracking. However, in the second and third line we replace n_0 by n'_0 which gives 2^{nd} order terms in l_s and m_s . Ignoring higher order terms we get

$$\begin{aligned} \varphi / 2\pi = & \delta l'_0 U' + (m'_0 + \delta m'_0) V' + n''_0 W' \\ & + l'_s U'_0 + m'_s V'_0 \\ & - (l'^2_s + m'^2_s) W'_0 / 2 \\ & - (l'_s \delta l'_0 / n'_0 + m'_s (m'_0 + \delta m'_0) / n'^2_0) W'_0 / 2 \end{aligned} \quad (3.54)$$

where we introduced $U'_0 = (U' - \delta l'_0 W'_0)$ and $V'_0 = (V' - (m'_0 + \delta m'_0) W'_0)$ and $W'_0 = W' / n'_0$. Equation (3.54) shows in the first place that we need also fringe tracking for $\delta l'_0$, $m'_0 + \delta m'_0$, and n'_0 (that defines a position on the unit sphere) to place the centre of each instantaneous snapshot image at $(l'_0 = 0, m'_0)$. In the second place the coordinates U'_0 and V'_0 are made “sky tracking” just by updating $\delta l'_0$ and $\delta m'_0$ as functions of time. In the third place it shows that all quadratic terms in l'_s and m'_s could be combined, which allows correction with a quasi-convolution as described in section 3.4. Finally, the last error term in the fourth line of (3.54) has cross terms that evaluate as

$$\delta\varphi_c / 2\pi = m'_s (l'_s \delta l'_0 + m'_s \delta m'_0) (m'_0 / n'^2_0) W'_0 \quad (3.55)$$

After correction for first and second order terms in l'_s and m'_s in (3.54) we are left with a 3^{rd} order phase error term given by (3.55) where l'_s , m'_s , $\delta l'_0$ and $\delta m'_0$ are small and by even smaller 4^{th} and higher order terms that have been ignored.

3.5.3 Duration of a synthesized snapshot observation

For a tracking interval $\Delta l'$ centred at $(l_0 = 0, m_0)$ we get $-|\Delta l'|/2 < |\delta l'_0| < |\Delta l'|/2$ and we have almost the second situation described by (3.50a) in subsection 3.4.5 for the maximum FoV in l_s and m_s taking $|\Delta l'|/2$ as the angular radius of that FoV. Actually we have some freedom to select $|\Delta l'|$ differently, and this defines a third situation with a maximum undistorted FoV for which 2^{nd} order terms need to be corrected by a quasi-convolution. This third type of FoV is defined by the tolerance for the 3^{rd} order tracking term and could in principle be smaller than the station beam, requiring faceting.

We now define the radius of the distortion free FoV by $\Delta\theta_{\text{cr}}$ for this third situation by using (3.55) with $|l'_s| \sim \Delta\theta_{\text{cr}}$ and $|m'_s| \sim \Delta\theta_{\text{cr}} \cos \theta_0$ for $\delta\varphi_c = \pi^{-1}$ while $m'_0 = \sin \theta_0$ and

$n'_0 = \cos \theta_0$ for zenith angle θ_0 . We neglect the term with $\delta m'_0$ and use $W = H / \lambda$ to get the proper order of magnitude for the tracking range

$$|\Delta l'| \sim (n'_0 / \pi \Delta \theta_{cr})^2 (\lambda / H) / m'_0 \quad (3.56)$$

Close to the Zenith we get a large tracking interval due to small m'_0 while $n'_0 \sim 1$ but at elevation below 45° the tracking interval reduces rapidly since $m'_0 \sim 1$ but n'_0 becomes small. This result is different from (3.50a) where we assumed $|\Delta l'| \sim \Delta \theta_{cr}$.

In the synthesized snapshot, the maximum phase error of π^{-1} rad appears in the visibility on the baseline with the largest non-planarity for a source at a specific position at begin and end of the tracking interval. Averaged over the interval we find a degradation of this visibility given by $\text{sinc}(\pi^{-1}) \sim (1-0.017)$, which is the same value as caused by the averaging over integration time and over bandwidth, and considered acceptable.

If the maximum tolerated phase error of π^{-1} rad in the visibility is on the longest baseline, for a source at the edge of the field this corresponds to a position shift of at most $\sim 0.3 w / 2\pi \sim 0.05 w$, where w is the half power beam width of the synthesized beam. However, the average position shift over the synthesized snapshot is zero and only a small broadening will appear.

According to table 3.1, Earth curvature gives at distance $L = 80$ km from the centre of the array $H = 500$ m. Our worst case LOFAR situation has at 50 MHz for an LBA station with effective diameter $D = 32$ m a beam diameter $1.2 \lambda / D \sim 0.22$ rad FWHM. Requiring a single FoV with radius $\Delta \theta_{cr} = 0.11$ rad gives at zenith angle 45° a maximum tracking interval $\Delta l' = 0.07$ or about 16 min for a source at meridian transit. For example, at 30° elevation, which is a practical lower limit, the tracking time is ~ 7 min. These examples show that synthesized snapshot imaging is an option in principle for the worst case Dutch LOFAR configuration that avoids faceting but then needs ~ 40 snapshot images for a synthesis of 6 h duration. As will be shown in section 3.7, such a number of Fourier images involves less processing than other processing steps like convolution and source subtraction, and makes the synthesized snapshot approach feasible for continuum observations. Moreover, these images can now be corrected in the image domain for image artefacts like field rotation, polarization rotation and changing beam shape that will be discussed in the next subsection.

3.5.4 Field rotation during sky tracking

Tracking a sky object with an array on a rotating Earth involves shifting in $\delta l'_0$ only for a field that is in culmination at the celestial equator. In all other cases a simultaneous shift in $\delta m'_0$ has to be made as well as a field de-rotation that compensates

for the change in parallactic angle. The shift operations could be handled by fringe tracking and by tracking of the U'_0, V'_0 -coordinates which together create a proper image in l'_s, m'_s -coordinates centred at l_0, m_0 pointed at a proper sky position for the middle of the short synthesis interval.

If the array were located at one of the Earth poles, then a simple rotation of its U, V -coordinates would give proper tracking for a field located at the same pole. For a field closer to the equator we need to make a shift correction to (l'_0, m'_0) and we choose $l'_0 = 0$ for the centre of the observation to get a FoV that can be extended by a quasi-convolution correction as discussed in the previous sections. Making rotation corrections to the U'_0, V'_0 -coordinates, will give an exact de-rotation of individual snapshots relative to the centre of the field at l_0, m_0 during the tracking interval given by $-\Delta l'/2 < \delta l'_0 < \Delta l'/2$.

For an array at lower latitude than the pole we first need a tilt of the l, m -plane towards the equator plane before the rotation correction can be applied, requiring a simple scaling in U'_0, V'_0 -coordinates. After rotation we need to tilt back the rotated U'_0, V'_0 -coordinates by rescaling. Such a tilt correction allows a rotation correction that is correct only for the centre of the field and approximately correct for the near vicinity. We need only a small rotation correction that the difference in parallactic angle for the centre of the synthesized snapshot and for the constituting snapshots.

Averaging of two de-rotated snapshot images that are observed symmetrically to (l'_0, m'_0) leaves no average shift in l'_s, m'_s -coordinates but only a small broadening of objects proportional to distance from the field centre. Using the associated decorrelation per baseline shows that this broadening can be ignored for the planned field sizes.

The larger parallactic rotation between the coordinate systems of the different synthesized snapshot images has to be taken into account when these are combined in a full synthesis image and needs some more explanation. When a field centre follows a sky track close to the Zenith we could expect according to figure 3.5 a fast change in orientation of the l'_0 -axis between successive short synthesis observations. The angle in the l, m -plane between the directions from field centre towards projected Zenith Z and projected Celestial Pole CP respectively is related to the parallactic angle defined for the great circles on the sphere and will also change rapidly. This effect is the consequence of our particular choice for the reference frame of a synthesized snapshot. However, for each individual snapshot in the chosen coordinate system we have only a slow rotation and a slow shift of the sky field as determined by the distance between the field and the Celestial Pole, which is corrected by rotation of the U'_0, V'_0 -coordinates as discussed above.

However, the orientation of the polarization vector as observed by antennas aligned along l - and m -axis respectively is aligned to a reference axis in the l, m -plane that is defined by the processing of the antenna signals as will be discussed in section

3.6.2. This means that the polarization orientation of a tracked field rotates only slowly relative to the field image. Averaging a rotating polarization vector gives a degradation factor for the linear polarization of $\text{sinc}(\delta/2) \sim 1 - \delta^2/24$ where δ equals the total rotation during the short synthesis. The latter is determined by Earth rotation of 0.0044 rad/min and after 10 min the reduction in linear polarization power is less than $8 \cdot 10^{-5}$, which can be ignored.

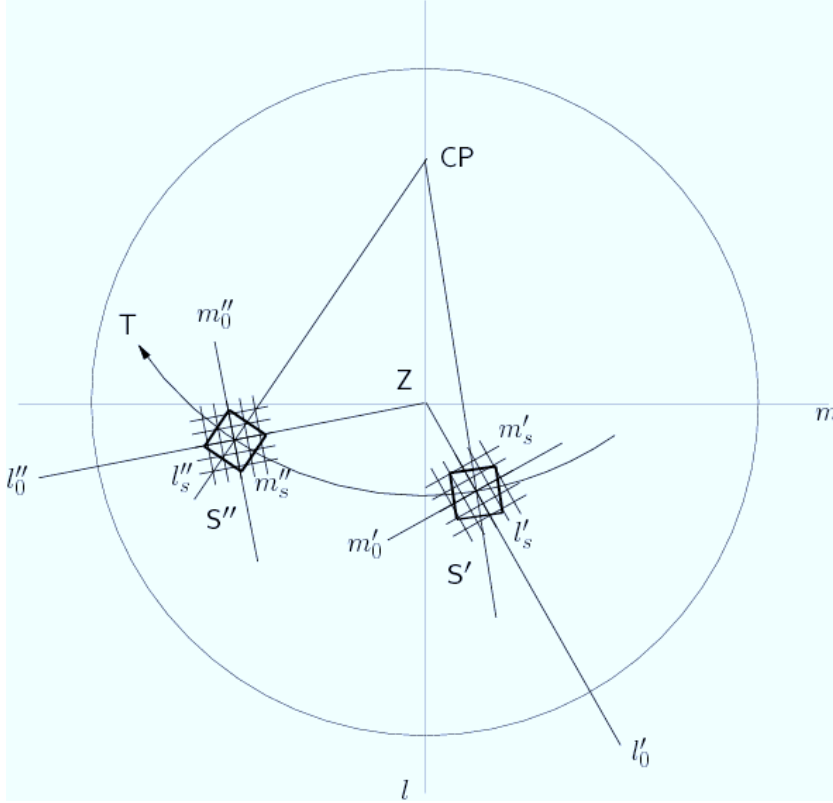


Figure 3.5. Spherical projection of source field S that follows a track T around the Celestial Pole CP in the l, m -coordinates of the reference system for array and stations. Snapshot images are visualized for the middle of the short synthesis intervals at (l'_0, m'_0) and (l''_0, m''_0) . The sky field S that is oriented towards the CP appears as squares S' and S'' that are rotated relative to the l'_s, m'_s - and l''_s, m''_s -coordinate systems for each short synthesis interval respectively.

When synthesized snapshot images are to be combined in a coordinate system with reference axis towards the Celestial pole we need to correct the coordinates of the synthesized snapshot images for some projected parallactic rotation angle. Apart from the coordinate conversion also the direction of polarization needs to be adapted to the convention for the new coordinate system. In the synthesized snapshot case we need correction for the projection of an angle between the great circle

direction from image centre towards Celestial pole and the polarization reference axis in the l,m -plane. This polarization rotation is realized by proper weighting of the four polarized visibility signals from the cross-correlation as will be discussed in section 3.6.2.

Also ionospheric Faraday rotation needs correction and appropriate time scales for update of this correction, which will be discussed in chapter 4.

3.5.5 Synthesis imaging with synthesized snapshots

A synthesized snapshot image is the sum of a series of instantaneous snapshot images, where the fringe stopping performs a continuous 3-D shifting operation on the visibilities. In fact we have a rotation in 3-D space over the unit sphere of the adopted reference coordinate system. However, 2-D Fourier imaging of a non-planar array in that adopted coordinate system gives phase errors of which the 2nd order ones (in field coordinates) can be corrected by a quasi-convolution operation on the instantaneous U,V -coordinates. However, 3rd order terms limit the FoV and the tracking range within the adopted reference coordinate system. The instantaneous U,V -coordinates for each snapshot image are convolved to a rectangular grid and corrected for scale, for position shift and for rotation of the centre of each instantaneous field. A FFT produces a synthesized snapshot image in rectangular l,m -coordinates centred at (l'_0, m'_0) for the centre of the tracking interval and the applied corrections per instantaneous snapshot image are approximately correct for a limited FoV. For objects at larger distance from the centre we find increasing phase errors on baselines with a large non-planarity with a maximum of ~ 0.3 rad at the edge of the FoV. This result is valid in any coordinate system, but in the special case of a reference system with W' -axis toward Zenith at the centre of the array we have small non-planarity due to Earth curvature, which allows a large FoV for long baselines.

Long synthesis observations take 2 - 24 h and a synthesis image is made as sum of a series of synthesized snapshot images, each < 10 min. For each synthesized snapshot we define a new and slightly rotated coordinate system with its U' -axis towards the position $l'_0 + \Delta l'$ where our field of interest will be at the middle of the next short synthesis interval. The pixels at $(l'_s + l'_0, m'_s + m'_0)$ have for each synthesized snapshot image a different conversion to coordinates that are fixed on the sky sphere.

First we need appropriate correction of each synthesized snapshot image for the polarized beam shape that is defined in the l,m -coordinate system of the stations, followed by interpolation on an appropriate sky bound coordinate grid that is defined for the nominal field position. Such an interpolation requires not only rescaling of the individual fields in l'_s, m'_s -coordinates for the change in m'_0 , but also for the change in parallactic rotation of the coordinate grid defined for the middle of the short syn-

thesis observation. Rescaling is more than a single scale factor determined by the elevation of the centre of the field, but includes also a varying scale factor over the field. Rotation and rescaling are combined if a 3-D rotation is performed from l', m', n' -image coordinates to l, m, n -sky coordinates and will only need an interpolation kernel with limited extent. In figure 3.6 we visualize this in two instances S' and S'' of the source field S at positions (l'_0, m'_0) and (l''_0, m''_0) respectively along the track T in the reference l, m -coordinate system.

The interpolation could for instance be realized by a convolution kernel, which in principle also allows additional corrections for field distortions such as differential refraction. A serious effect is that the interpolation results into some change in image scale relative to the centre of the image not only for sources, but also for side lobe responses. The result is that in the new coordinates the psf of a source is no longer position invariant as in de FFT coordinates. The consequence is that image deconvolution with a fixed psf pattern for different places in the image has limited accuracy. It means that side lobes that emanate from sources outside the FoV of a facet image have to be eliminated from the visibility data by subtraction of the emanating source according to some model.

Adding synthesized snapshots with different scaling and rotation corrections produces an image where the sources add at their nominal locations. The psf pattern of each synthesized snapshot, defined by the fixed baseline configuration of the array, is scaled, and rotated before adding them together. As a result the final side lobe pattern in a long synthesis is scrambled and allows a simplified estimation of its rms value, which will be used in chapter 5.

Defining a convenient sky coordinate system that simplifies the interpolation of all facets of a series of short synthesis image grids on a sky grid is outside the scope of this dissertation. For estimating the processing needed by interpolation of the synthesized snapshot images we just assume that each point in each snapshot image is interpolated to 4×4 pixels on a grid in the output image. This size is comparable to the extent of the main lobe of the synthesized beam and must be sufficient for accurate interpolation. Although other schemes are possible as well we will use in section 3.7 the required processing of such 4×4 image interpolation as first order estimate for comparison with the processing required for visibility convolution and for Fourier transformation.

3.5.6 How do sources outside the nominal FoV appear in a synthesised snapshot image?

In fact, this question was one of the drivers to investigate wide field synthesis imaging aspects, since the LOFAR stations have less suppression for sources outside the main beam than filled apertures. Moreover, LOFAR operates in a frequency band shared with terrestrial transmitters, luckily mostly located near the horizon of

the array. An important question is how such sources appear in an Earth rotation synthesis observation, which depends on the various approximations made in the imaging process. Even more important is how the side lobes of all sources outside the main beam of the stations contribute to the noise level of an approximate synthesis image that satisfies certain tolerances only in a small section of the sky.

LOFAR has its antennas close to the ground, giving a local horizon at ~ 4 km distance around the LOFAR core. This area has no Radio Frequency Interference (RFI) sources higher than the LOFAR antennas. Potential RFI sources outside the LOFAR core area have limited height and are observed at very low elevations where the LOFAR antennas have low sensitivity. However, satellites have higher elevation and even wings of airplanes and windmills produce observable reflections of distant transmitters.

An all sky Fourier image with a planar phased array station will therefore show only few RFI sources, most of them close to the horizon [Wijnholds, 2004]. Signals generated by intermodulation of two strong monochromatic point sources in the receiver chain of each phased array element show up as point sources at a different frequency and different position that can be predicted [Boonstra, 2005]. The same is true for snapshot imaging with the planar core array, but objects closer to the core than $4L_c^2/\lambda$ will appear blurred, where L_c equals the core radius.

In a series of snapshot images, the sky sources appear to move all differently since they are located on a sphere that rotates around the polar axis. This gives in snapshot imaging different projections on the horizontal plane as function of time. Earth rotation synthesis imaging combines the snapshots in a way that corrects this sky movement approximately for a limited FoV. This issue has been addressed in subsection 3.1.10 for conventional imaging and in subsection 3.5.4 for snapshot imaging and we will analyse the consequences for imaging of sources outside this FoV, such as for distant RFI sources.

Sources with a fixed position relative to the array have a constant geometrical phase. A point source on the pole also has a constant geometrical phase and we could therefore expect that all sources with constant geometrical phase will appear at the pole in Earth rotation synthesis. This reasoning is too simplistic since it ignores (i) the actual phase values for each interferometer and (ii) ignores the de-rotation in the imaging process to compensate Earth rotation. We will therefore look more into the details of the snapshot imaging process. In subsection 3.5.6.1 we start analysing the attenuation of signals by the side lobes of a station. In subsection 3.5.6.2 we will discuss the snapshot imaging approach with phased array stations and in subsection 3.5.6.3 we summarize our results.

3.5.6.1 Attenuation by side lobes of a phased array station

As will be shown in subsection 3.6.1, a phased array station has a phase reference position where signals from different directions have the same phase as if a point like antenna was used at that position. The beam formed signal received through a side lobe has this same phase although the amplitude is the sum of signals from the different elements. As a result, only the amplitude is reduced since the element signals all have a geometrically different phase depending on the direction of the source. This produces the side lobes of the station beam, and is the result of the same principle as explained in subsection 3.2.1 for bandwidth decorrelation leading to (3.28). This analysis assumes however a uniform signal distribution, like the one provided by the illuminated aperture of a dish. This analysis is for the LOFAR LBA stations only approximately valid, since the antenna elements have a non-uniform distribution.

The LOFAR stations are rotated w.r.t. each other, which means that the two stations of an interferometer could observe a source each through a different side lobe [Bregman, 2012]. The voltage pattern of the station beam has alternating positive and negative side lobes. This leads for an interferometer between two rotated stations to alternating sign flips when a source moves through the side lobes. These alternations occur for different baselines at different instants. As a result, a point source outside the main beam will show up distorted and attenuated in addition to the attenuation by the side lobes of the stations [Wijnholds, 2008]. Only objects that are strong enough to have sufficient SNR per baselines are self-calibrated and subtracted accurately for each period where no sign flips occurs. All weaker objects outside the main beam suffer from sign flips in their visibilities leading to blurred and attenuated structures. As a result, their side lobes are reduced as well, but contribute to the side lobe noise in a synthesis image, which is the subject of chapter 5.

This LOFAR situation is different from the WSRT where all dishes are identical and all antenna patterns are aligned within a fraction of the width of a side lobe beam. Moreover, the equatorial mount of the WSRT telescopes keeps sky objects in the same side lobes when a sky field is tracked. However, sources at a fixed position relative to the telescopes are scanned when the main beam tracks a source field. For LOFAR, which has in fact an alt-azimuth mount, all sources outside the main beam will be scanned, and therefore the visibility amplitudes vary with the amplitude of the side lobes. In addition, there is also a decorrelation effect by bandwidth and integration time in processing of the correlated visibilities.

3.5.6.2 Rotation and fringe track effects in synthesized snapshot imaging

In snapshot imaging we have a Cartesian U, V, W -coordinate system with direction cosines l , m and n attached to the array with W -axis towards local Zenith, U -axis and V -axis are in the horizontal plane, and the V -axis towards the azimuth of the

source field of interest. 2-D Fourier inversion of the visibilities in the U, V -plane provides an image of the hemisphere above the local horizon. In practice, a 3-D fringe shift is performed and only a small image is made centred at the position where the main beam of the telescopes is pointed. For a planar array with $W = 0$ the 3-D fringe shift is equal to a 2-D fringe shift in the horizontal l, m -plane as discussed in subsection 3.1.6.

The observed sky sphere above the horizon rotates around the polar axis and its projection on the horizontal plane changes shape that cannot be corrected by only a shift in the l, m -plane but needs in addition a rotation correction as discussed in subsections 3.1.5 and 3.5.4. The source vectors in l, m, n -space of the snapshot image need to be rotated to l', m', n' -space of the celestial sphere before snapshot images can be co-added to a synthesis image [Wijnholds, 2005]. In this process where sources have fixed positions in the l, m -image plane, they become smeared along tracks on the l', m', n' -sphere, while moving signals from a satellite or from reflection by an airplane just get a track at a different location and with a different orientation. As a result the average intensity at a position along the track is reduced compared to the intensity of sky sources that are integrated on a fixed l', m', n' -position. For a point like source this attenuation factor is just the resolution width divided by the length of the source track.

In practice we have to deal with fringe tracking that makes a particular sky location stationary in position. This position will then be the centre of a 2-D Fourier image, but de-rotation of the sky field depends on the particular imaging approach.

We have shown in subsection 3.5.4 that the first-order de-rotation correction in a synthesized snapshot leads only to small smearing of sources within the FoV that can be ignored. For sources at larger distances, this leads to short tracks. We have shown in subsection 3.5.5 that the synthesized snapshot images can be fully corrected for the field distortion associated by the varying parallactic rotation over the field. However, when large synthesized snapshot images would be made, distant point sources will get longer tracks. Combining such longer tracks, results in blur of these sources in the combined image.

Another effect of the fringe tracking is that other positions, such as of a RFI source at the horizon, get an additional fringe rate that leads to decorrelation by averaging to samples with finite extent in frequency and time domain as discussed in section 3.2. As a result, sources that appear imaged as a sky track are further reduced in intensity. In a small image that covers only the field of interest on the sky, we only suffer from the side lobe responses of this source track. The value of such side lobe contributions in an image will be discussed in chapter 5.

3.5.6.3 Summary and conclusions

We have shown how a source with fixed position relative to the array will show up as a track in a large synthesised snapshot image where each individual snapshot

with a planar 2-D array is only corrected for sky rotation to first order. The side lobe responses of this track could cause observable structure in a smaller synthesis image.

We have identified six attenuation mechanisms that determine the effective strength of this track.

- The first one is the attenuation by the element pattern of the antennas in a phased array station, which has particularly low values near the horizon.
- The most important attenuation is by the side lobes of the station array beam, which can be controlled by appropriate tapering of the station array.
- Fringe tracking for a particular sky direction creates a fringe rate for all signals from objects at different locations, which leads to attenuation of visibility signals from these directions by integration over time.
- The fourth attenuation mechanism is the result of the sign alternations in interferometer visibilities when the sky sources are scanned with different side lobes of the different station combinations.
- Sources at large distances from the main beam get only partial correction for Earth rotation in a synthesized snapshot image, which leads to short and attenuated tracks.
- In a small synthesis image, we suffer from the side lobes of these tracks.
- Very wide-field synthesized snapshot images with different parallactic angle have different track orientations. Combining such images will lead to blur of sources at large distance from the field centre, which is a sixth attenuation mechanism.

A complete overview of possible internal generated interference as well as calibration and imaging artefacts is outside the scope of this dissertation, but we have shown that the impact of sources outside the station main beam is low for Earth rotation snapshot synthesis imaging with a planar array. Quasi-planar arrays, like LOFAR, suffer from additional blur in sources at large distance from the field centre. The impact of such blur on the average side lobe level is however small as will be discussed in chapter 5.

3.5.7 Summary and Results

The conclusions of the different subsections can be summarized as follows:

- The visibility phase of a point source observed by a quasi-planar array can be described by a fringe shift term, a 2-D Fourier kernel for projected baselines and deviation terms. These deviating terms are proportional to the distance from the fringe centre, and to the non-planarity of the stations in the array. The difference in Z-coordinate of the stations in the chosen co-

ordinate system for 2-D Fourier imaging defines the non-planarity of a baseline.

- In this case, the baselines are projected on the reference plane chosen for the 2-D Fourier transform from a direction parallel to the direction of the field centre as depicted in figure 3.3.
- The phase deviations as function of position in the field and as function of baseline cause deviations in the shape of a point source that depend on its position in the field. Therefore, also the side lobes of this point source appear distorted.
- This description forms the basis for 2-D Fourier snapshot imaging, where only a small image is made that is centred on the so-called fringe tracking centre and where point sources suffer from deviations that increase with distance from the centre of the image.
- By appropriate choice of a U', V', W' -coordinate system there is, after appropriate projection of the baselines on the U', V' -plane, a dominant phase deviation term per baseline that is quadratic in the distance from the field centre and proportional to the non-planarity W' of the baseline. It can, however, be corrected by a complex convolution as introduced in section 3.4.
- In subsection 3.5.2 we have shown that after correction of the phase terms that are quadratic in the distance from the centre of the image. The residual phase deviation is dominated by third order terms that are proportional to products of image coordinates l'_s and m'_s and proportional to distances $\delta l'_0$ or $\delta m'_0$ between actual position and nominal centre position of the image.
- We introduced the concept of a synthesized snapshot observation in subsection 3.5.3, where the U', V' -coordinates of each of the constituting instantaneous snapshots are not only corrected for a projected W' -term but also need a differential field rotation correction relative to the middle of the observation interval as discussed in subsection 3.5.4.
- Synthesized snapshot observations shorter than about 10 min do not need such relative rotation correction for the polarization orientation during the observation since the average intensity of rotating polarization is only reduced by 0.008 %. However, each synthesized snapshot needs appropriate parallactic image rotation and Faraday rotation correction before adding with other synthesized snapshots.
- Each synthesized snapshot image with $l'_0 = 0$ and $m'_0 = \sin(\theta_c)$ has a position invariant psf in l'_s, m'_s -coordinates where θ_c is the zenith angle of the centre of the short synthesis image at the middle of the tracking interval. All conclusions thus far are generic for any choice of the coordinate system.
- For a quasi-planar array such as LOFAR, where the stations follow the Earth curvature, we define a U, V -plane by a best fit to the plane of the ar-

ray. A maximum residual third order phase deviation of π^{-1} rad defines the maximum tracking range for a given maximum FoV and a given maximum non-planarity. We define a FoV extending to half power in the beam of a LBA station with 32 m effective diameter at 50 MHz pointed at Zenith. The maximum phase deviation for baselines between core and stations at 80 km distance is reached at begin and end of a 16 min tracking interval for an elevation of 45° . At 30° elevation the tracking interval is ~ 7 min and leads to at most 1.7 % signal decorrelation on these baselines for objects at the edge of the FoV.

- For stations at larger distances faceting is needed to keep tracking time and size of the complex convolution kernel at low values to make processing affordable, which will be discussed in subsection 3.7.
- A typical 6 h synthesis requires less than 40 synthesized 2-D Fourier snapshot images with varying parallactic rotation which requires proper scaling, rotation and interpolation on a sky grid before averaging to a full synthesis image.
- Also the polarization orientation needs to be corrected by proper conversion of the four observed polarization coherencies. The required rotation is in this case given by the projection of a rotation angle between the great circle direction from image centre towards Celestial pole and the polarization reference axis in the plane of the array.
- The result of the interpolation and rescaling to sky coordinates is that the position invariant psf in the coordinates of the FFT images is replaced by a position dependent one in the new coordinate grid, which complicates deconvolution procedures that use iterative subtraction by an assumed position invariant psf in the image domain.

The main result of this section is a new Earth rotation synthesis imaging procedure that forms a single large image by interpolation of individually corrected synthesized snapshot images:

- Each synthesized snapshot image needs only a small complex convolution kernel that corrects primarily for second order effects by the intrinsic non-planarity of an array where the stations follow Earth curvature.
- Projection from the direction of the source on the horizontal plane of the array causes third order effects that limit the duration of a synthesized snapshot.
- This limited duration is matched to inaccuracies that arise by the small differential parallactic rotation corrections that need to be made to the projected U,V-coordinates of the observed visibility samples.
- The large field correction for parallactic rotation between synthesized snapshots can be combined with changing refraction over the wide field when the synthesized snapshot images are combined to an image in sky coordinates.

- Sources outside the FoV of a synthesized snapshot image will be smeared to short tracks since the first-order correction for parallactic rotation is incomplete.
- Side lobes of these short tracks will increase the noise level in an image.
- An attractive aspect of the method is that other corrections as function of location in each synthesized snapshot image, such as for beam polarization, parallactic polarization rotation and global Faraday rotation can all be applied conveniently in the image domain.

3.6 Phased array station beam aspects in synthesis imaging

In this section we indicate how image formation with a synthesis array is impaired by the introduction of assumptions that have been made to derive (3.8) from (3.6), which is the basis for Fourier inversion according to (3.14). A very important aspect of the station beam is that it needs sufficiently low side lobes, strongly reduced grating lobes and sky-tracking with sufficient precision. These requirements reduce the processing capacity needed for imaging as will be discussed in chapter 5, but more importantly support self-calibratability [Wijnholds, 2011] by limiting contribution of sources outside the main beam which will be discussed in chapter 4. The focus in this section is on the impact of beam shape and beam polarization on synthesis imaging based on summation of short synthesis images that are individually corrected. These station beam and polarization corrections are independent of the reference frame in which synthesis images are made. The maximum duration of a synthesized snapshot image should according to the derivation in section 3.5.4 be of the order 10 min and avoids correction for rotation of the polarization during this period. We will check whether the change of the beam pattern in shape and polarization during such a tracking interval needs a faster update rate.

An important assumption in deriving (3.14) is that all interferometers have the same FoV beam g_k^p as defined by the product $g_{ik} g_{jk}^*$ of the two station voltage beams. However, the LOFAR stations are phased arrays and are not identical by design in an attempt to reduce the side lobes and especially the grating lobes of the averaged station beam patterns. Although the same pattern for the array configuration of all stations is used, each station has a different pattern orientation and consequently a different beam pattern on the sky [Wijnholds, 2008].

In the following paragraphs all beam effects that have an impact on synthesis imaging will be introduced and a proper context is provided by describing how these effects are used or mitigated in the actual design of LOFAR.

In case station beams are different we have the situation that for a given off-axis point source the observed visibilities are not equal for all baselines. The result is

that the Fourier transform (3.14) gives distortions to each point source depending on its location. The shape of a point source is no longer equal to the nominal psf of the Fourier transform that is based on the distribution of the U, V -samples and their nominal weights. However, the peak of the psf is just the average of all visibilities and therefore defines the intensity at a specific position as the average station beam of the synthesis observation. By the same argument we can therefore simply estimate the beam g_k^p in (3.14) by forming a weighted average of all products of station voltage beams that are used in the Fourier transformation. This weighted average over all baselines predicts the proper attenuation for each point source in the field. However, extended objects that are resolved on long baselines have no contributions from stations that contribute to these long baselines. Consequently the required weighing scheme of station beams becomes dependent on source structure. Only for identical station beams is the effective weighting of visibility signals independent of direction within the average beam of all stations.

We can describe the station beam as the product of an element antenna pattern and the array pattern beams of the station. This beam product description would be correct if Electro-Magnetic (EM) interaction between the elements could be ignored, as is usually the case for interaction between stations. The element antennas in the LOFAR stations are however so close to each other that EM interaction cannot be ignored and results in two effects. One effect is that the antenna beam of each element is distorted compared with the pattern of a free standing element. The result is that an incident plane wave induces in each element a different voltage depending on direction of arrival. The second effect is that a current in one element induces voltages in all other elements that are connected by the so called mutual impedances. An array with N elements has an $N \times N$ impedance matrix that determines the current in the load impedance attached to each element. A tedious EM simulation is required to determine all the patterns and the impedance matrix of an array for a large set of frequencies [Cappellen, 2006]. Then for each specific direction a separate array pattern with different side lobe structure has to be calculated since the element antenna patterns are all different. In fact we have the same problem as described before as for the synthesis image where the side lobe structure of point sources also varies with direction.

There are two methods to deal with beam problems associated with varying beam shape, one by reducing the effects in each station and one by reducing the effects in a synthesis observation, while both can also be combined.

The LOFAR High Band Array has its element antennas on a regular grid, which results in grating lobes at the higher frequencies where the wavelength is shorter than twice the element separation. EM coupling between antennas on a regular grid creates not only strong fine structure in individual antenna beams, but also a specific structure in the elements of the complex mutual impedance matrix. The latter effect results in so-called blind angles where the average antenna pattern of all elements has strongly reduced sensitivity for specific directions that depend on

frequency. A simplified method has been developed allowing first order estimation of the blind angle effect [Wijnholds, 2008] and will be further discussed in section 3.6.3. Grating and coupling problems are mitigated in LOFAR by using a different orientation of the station geometry for each station. In the beam of each interferometer the grating lobe of one station is multiplied with a low side lobe of the other station, which results in small remaining grating lobes that are the geometric mean of a large and a very small lobe. Since blind angles in a station beam give typically less than 50% reduction in sensitivity, the geometric mean with the full intensity pattern of the other station only halves the effect. Averaging over all interferometers where the remaining blind angles and grating lobes appear at different locations, leads to further reduction [Wijnholds, 2008]. The configuration of the LOFAR array has rings of stations around a centre location and the rotation of the stations is organized per ring, such that for all baseline ranges a reasonable mitigation occurs [Bregman, 2011].

For the LOFAR Low Band Array the effects are not only reduced by rotating the station configuration but also by using an element configuration with randomly varying separation between the elements. The main reason for such a randomized configuration is that grating lobes that would arise in a sparse regular array are now scrambled since the phases of the signals from the grating direction are randomized. The phases of EM interaction terms between a reference element and all other antenna elements are randomized as well and make deviations between individual element beams less pronounced. Finally, also the phases of the coupling impedances are randomized. The result is that the beam product description for an LBA station is indeed a reasonable approximation if the average of all different element patterns in the array is used as the effective element pattern [Cappellen, 2006].

It must however be realized that the voltage beam pattern of a phased array station could have direction dependent phase structure, which will be discussed in subsection 3.6.1.

The station array beam as produced by phasing signals in the beam former is scalar and has no polarizing characteristics itself, but the element antenna beams have strong polarization structures. This apparent polarization is related to the projection of the beam patterns of two orthogonal dipoles on the sky where the dipoles appear no longer orthogonal. Therefore, spurious polarisation is produced not only from field rotation relative to the dipole orientation but also by the movement of the station beam through the polarized pattern of the average beam of the element antennas when it tracks a sky source. To reduce the problems in synthesis imaging associated with station beams that have different polarization characteristics it has been decided that all antenna elements of all stations should have the same orientation on the sky. Therefore, all antenna elements in a station are counter rotated with respect to the station configuration rotation such that all dipoles in the core of the array have the same orientation and that the elements in all other stations are oriented as parallel as possible to the dipoles in the core. In that case the observed

sky can be described by a true brightness distribution multiplied by an average element beam with only a global polarization characteristic that is the same for all interferometers. Differences between beams of stations with different longitude and latitude need to be corrected together with differences in local Faraday rotation.

A station main beam that tracks a sky source from a rotating Earth suffers from a number of effects that change its shape and polarization characteristics:

- Elongated beam shape in elevation direction by foreshortening at larger zenith angles.
- Rotation of a sky field relative to station beam and element beam.
- Changing polarization characteristic over the station beam as determined by its pointing direction relative to the polarization structure of the average element pattern.
- Changing beam shape when the array beam passes over a blind angle or other structure in the average element beam.
- A separate effect is related to electronic cross-talk between signals from the two orthogonal dipoles of each antenna. This effect is however direction independent and less than -60 dB, giving less than 0.1 % polarization, which has a circular component depending on the phase of the cross-talk.

We will discuss a few of these properties in some more detail such as the location of the phase centre of each station in subsection 3.6.1. In subsection 3.6.2 we introduce the polarization formalism and show the basic characteristics of the beam of dipole-like antennas as used in LOFAR. In subsection 3.6.3 we explain how the average element pattern in a phased array station determines the polarization characteristics of a station beam. In subsection 3.6.4 we discuss the polarization characteristics of a station main beam if calibration is performed on the XX and YY channels based on a single un-polarized source close to the centre of the beam.

In subsection 3.6.5 we give an order of magnitude for the expected distortion effects in the station main beam due to blind angles in the average beam of all element antennas in a station. In subsection 3.6.6 we explain why LOFAR uses the same element antenna orientation for all stations. Finally we discuss the effects of non-equal station beams in subsection 3.6.7 and conclude with a summary in subsection 3.6.8.

3.6.1 Phase centre position of a phased array station

The phase centre of an antenna is defined as the reference position from which spherical radiation appears to emanate in the transmit situation. For a dipole antenna above a ground plane it is the point from which the sum of signals from dipole and its reflected image effectively emanates. For the LBA antennas where the metallic reflector is smaller than a wavelength the phase centre lies below the ground

plane on a depth determined by the effective dielectric constant and the conductivity of the soil. Since both are influenced by the amount of water content the effective height of a LBA station varies over time and between stations [Arts, 2005].

For an array of such antennas at position \mathbf{r}_n from a reference position \mathbf{r}_0 the N signals S_0 are co-added with weight w_n and provide the signal $S(\mathbf{l})$ when steered to direction \mathbf{l}

$$S(\mathbf{l}) = S_0 \sum^N w_n \exp(-2\pi i \mathbf{l} \cdot (\mathbf{r}_n - \mathbf{r}_0) / \lambda) \quad (3.57)$$

where \mathbf{l} is the vector of direction cosines in a Cartesian coordinate system with z -axis towards local Zenith while λ is the wavelength.

The station is calibrated and fringe stopped such that at $\mathbf{l}_z = (0,0,1)$ all signals arrive in phase giving

$$S(\mathbf{l}_z) = S_0 \sum^N w_n$$

For direction \mathbf{l} we then get

$$S(\mathbf{l}) = S_0 \exp(2\pi i \mathbf{l} \cdot \mathbf{r}_0 / \lambda) \sum^N w_n \exp(-i \varphi_n)$$

With $\varphi_n = 2\pi \mathbf{l} \cdot \mathbf{r}_n / \lambda$ and for small φ_n we approximate the equation by

$$S(\mathbf{l}) = S_0 \exp(2\pi i \mathbf{l} \cdot \mathbf{r}_0 / \lambda) \sum^N w_n (1 - i \varphi_n)$$

We can now evaluate the imaginary and the real parts of $\sum^N w_n (1 - i \varphi_n)$ and determine the phase

$$\arg(\sum^N w_n (1 - i \varphi_n)) = -2\pi i \mathbf{l} \cdot \mathbf{r}_w / \lambda$$

where \mathbf{r}_w is the weighted average station position given by

$$\mathbf{r}_w = \sum^N w_n \mathbf{r}_n / \sum^N w_n \quad (3.58)$$

So, $\arg(S(\mathbf{l}) / S(\mathbf{l}_z))$ readily evaluates as

$$\arg(S(\mathbf{l}) / S(\mathbf{l}_z)) = 2\pi i \mathbf{l} \cdot (\mathbf{r}_0 - \mathbf{r}_w) / \lambda \quad (3.59)$$

This equation shows that the phase of the calibrated array signal is independent of \mathbf{l} only when $\mathbf{r}_w = \mathbf{r}_0$, i.e. \mathbf{r}_w is the phase centre of the array. The phase of a properly calibrated station array given by (3.57) could be considered as the phase term of the station voltage beam pattern if an arbitrary station reference position \mathbf{r}_0 is used instead of \mathbf{r}_w to evaluate the baseline vector \mathbf{U} of an array.

In practice this means that if one or more elements fail we need to give them effectively zero weight in the beamformer to reduce receiver noise, but more importantly the station position \mathbf{r}_w has changed. We need therefore to change the position of a phased array station in the calculation of the U,V-coordinates of any baseline used in Fourier imaging that involves the hampered station. For a station with diameter D we have an average element distance from the centre of $\sim D/4$ and with N elements the effective position will change by $\Delta r_w = D / 4N$ if one element fails. The maximum phase change is $\Delta\phi = 2 \pi l_h \Delta r_w / \lambda$ for an object at half power in the station beam at $l_h = 0.6 \lambda / D$, which results in $\Delta\phi = 0.3 \pi / N \sim 1/N$ irrespective of station diameter or wavelength. For LOFAR with $N \sim 48$ we find $\Delta\phi \sim 1^\circ$ and proportionally smaller errors for objects closer to the centre of the beam when a single element fails. Not only the beam pattern of the phased array station is changed requiring a different beam correction, but all objects in the field get different phase errors that will create different distortions in the side lobe patterns of all objects if the station position is not adapted.

Interestingly there is no need to change the reference position of the station as is used by the source tracking at station level or by the fringe tracking at correlation or by fringe shifting during imaging, which together define the centre of the Fourier image that needs to be imaged. The reason is that the beamforming at the station corrects the signal phases of all elements for the direction of the centre of the field. Fourier imaging is in fact beam forming for offset directions from the field centre and needs to correct for the average phase of all elements in this offset direction. Fourier imaging needs proper U,V-coordinates based on the positions of a phased array station averaged over the positions of elements that actually contribute including the weight of the taper function that is applied after calibration of the element signals.

3.6.2 Array element beam patterns and polarization characteristics

Polarization is a confusing matter where issues of geometry, electromagnetic properties, electronic gain and calibration come together. Although the foundations for treating these issues for synthesis arrays with dish telescopes are well known in principle [Hamaker, 1996] extensions are needed to handle large FoV as for phased array stations [Carozzi, 2009].

When a dish telescope with a dual polarized feed is pointed towards a source, two orthogonal field components are observed that are parallel to a plane perpendicular to the direction of propagation. When a sky source is tracked with an Earth bound telescope, the telescope main beam stays pointed towards the source, but rotates around the pointed direction. As a consequence also the response of the antenna pair could change depending on the polarization content of the source.

A phased array antenna station has a number of dual polarization antenna elements that have identical orientation. A LOFAR phased array station has receptor elements with orthogonal dipole like antennas aligned along X- and Y-axis respectively as depicted in figure 3.6.

The signals of all x-antennas are added by the x-beam-former and the signals of all y-antennas are added by the y-beam-former. Since there is very low crosstalk (< -60 dB) between the x- and y-signals paths the polarization of the summed x- and y-signals is determined by instrumental polarization characteristics that are averaged over all x- and y-antennas and their receiver chains respectively.

An excited antenna radiates a field in a specific direction that has at a large distance an electric field vector \mathbf{e} with only two orthogonal components e_θ and e_ϕ perpendicular to the propagation direction. EM simulation of a single antenna provides a power beam pattern $P(\theta, \phi)$ that is normalized (at $\theta = \phi = 0$) given by

$$P(\theta, \phi) = g_\theta(\theta, \phi) g_\theta(\theta, \phi)^* + g_\phi(\theta, \phi) g_\phi(\theta, \phi)^* \quad (3.60)$$

where $g_\theta(\theta, \phi)$ and $g_\phi(\theta, \phi)$ are the normalized voltage beam patterns for the two field components.

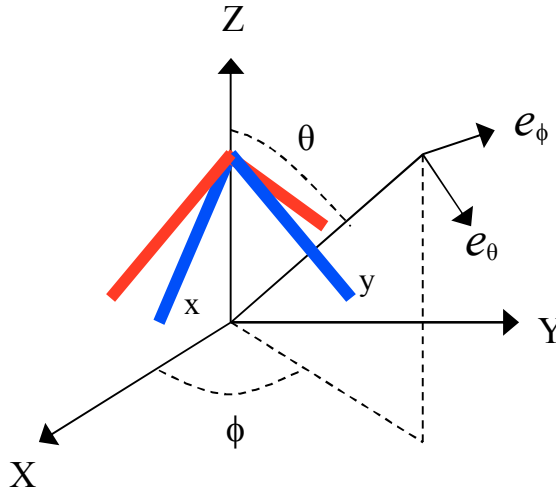


Figure 3.6. Antenna geometry in a phased array antenna station. Two dipole like antennas x and y are oriented along X- and Y-axis in the horizontal plane respectively with Z-axis toward Zenith (blue arms are the minus poles). A plane wave with electric field vector \mathbf{e} from a direction with zenith angle θ and azimuth ϕ has two orthogonal field components e_θ and e_ϕ .

In figure 3.7 we give as example the power pattern of a single LOFAR dipole like antenna at 80 MHz which shows enhanced beam width with highly elliptical shape due to its large height (in wavelength) above the ground plane.

On reception of a plane wave with field strength \mathbf{e} from direction (θ, ϕ) the voltage signal at the terminals of the x antenna is given by

$$v_x = g_{x\theta} e_\theta + g_{x\phi} e_\phi \quad (3.61)$$

This equation can be extended to a full matrix equation when we introduce the response v_y of a y-antenna that is ninety degrees rotated around the z-axis. Then v_x and v_y are the elements of column vector \mathbf{v} and the g terms form the so called 2x2 Jones matrix \mathbf{G}

$$\mathbf{v} = \mathbf{G} \mathbf{e} \quad (3.62)$$

The observed 2x2 coherency matrix \mathbf{V} is given by

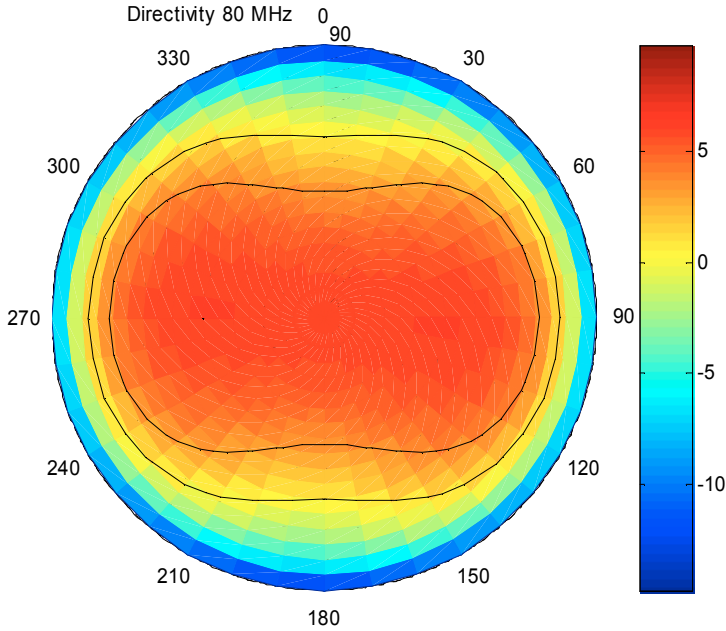


Figure 3.7. Elliptical antenna power pattern with -3 dB and -6 dB contours (relative to peak intensity) of a single LBA dipole at 80 MHz as function of azimuth and zenith angle (From [Arts, 2006]).

$$\mathbf{V} = \langle \mathbf{v} \mathbf{v}^H \rangle \quad (3.63)$$

where $\langle \rangle$ indicates a time average and H indicates the Hermitian transpose.

So (3.63) readily evaluates as

$$\mathbf{V} = \mathbf{G} \langle \mathbf{e} \mathbf{e}^H \rangle \mathbf{G}^H \quad (3.64)$$

where we assumed that the antenna beam patterns are constant over the short averaging period.

Inversion of (3.64) leads to

$$\mathbf{E} = \langle \mathbf{e} \mathbf{e}^H \rangle = \mathbf{G}^{-1} \mathbf{V} \mathbf{G}^{H-1} \quad (3.65)$$

where all elements of the matrices are a function of θ and ϕ .

The four coherence components of \mathbf{E} are related to the four Stokes parameters I, Q, U and V of an incident plane wave by [Hamaker, 1996]

$$\begin{aligned} E_{\theta\theta} &= \langle \mathbf{e}_\theta \mathbf{e}_\theta^* \rangle = (I + Q)/2 \\ E_{\phi\phi} &= \langle \mathbf{e}_\phi \mathbf{e}_\phi^* \rangle = (I - Q)/2 \\ E_{\theta\phi} &= \langle \mathbf{e}_\theta \mathbf{e}_\phi^* \rangle = (U + i V)/2 \\ E_{\phi\theta} &= \langle \mathbf{e}_\phi \mathbf{e}_\theta^* \rangle = (U - i V)/2 \end{aligned} \quad (3.66)$$

where * indicates complex conjugation.

In the same vein, we could define a set of four observed Stokes parameters based on the four observed coherence components of \mathbf{V}

$$\begin{aligned} I_V &= V_{XX} + V_{YY} \\ Q_V &= V_{XX} - V_{YY} \\ U_V &= V_{XY} + V_{YX} \\ i V_V &= V_{XY} - V_{YX} \end{aligned} \quad (3.67)$$

If we arrange the four true Stokes parameters in a column vector \mathbf{S} and construct an observed Stokes vector \mathbf{S}_V from the four observed coherencies we find the so called 4x4 Mueller matrix \mathbf{M} that relates the two Stokes vectors.

$$\mathbf{S} = \mathbf{M} \mathbf{S}_V \quad (3.67a)$$

Equation (3.67a) is the equivalent of (3.65) and M can be constructed from the two Jones matrices of the antennas that form an interferometer [Hamaker, 1996].

Combining figure 3.7 with a 90 degree rotated one we find for an un-polarized plane wave from direction (θ, ϕ) the power responses, where observed Stokes parameter

I_V has in azimuth an almost circular shape shown in figure 3.8 and observed Stokes parameter Q_V has a 2-fold symmetric shape shown in figure 3.9.

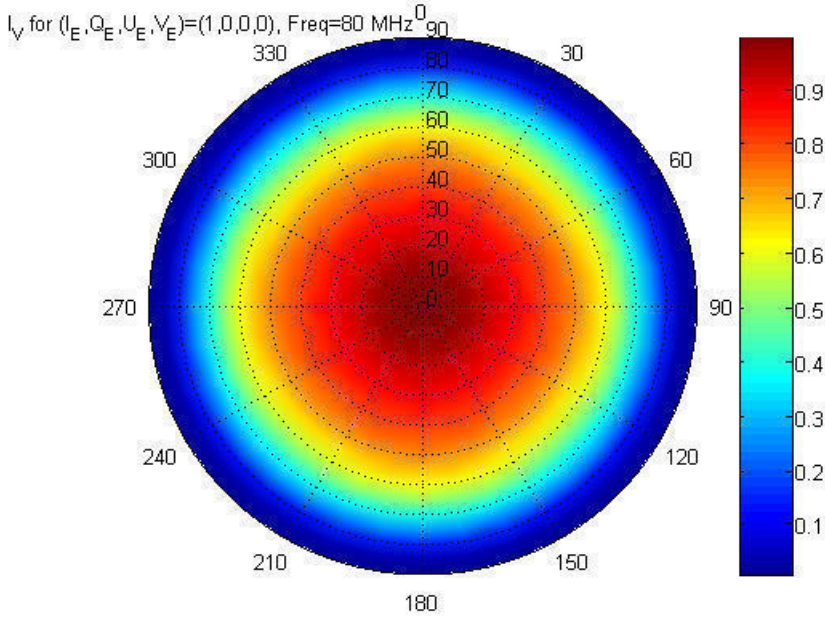


Figure 3.8. Almost circular antenna response pattern in total intensity I of a dual polarized LOFAR LBA antenna for an un-polarized input wave at 80 MHz as function of azimuth and zenith angle (From [Arts, 2005])

The typical cloverleaf pattern of the instrumental linear polarization is also found for the beam of a dish telescope but covers in that case the main beam till the first null.

It is important to realize that in electronic engineering the cross-polarization over a beam is expressed as a power ratio defined as $g_{x\phi}(\theta)g_{x\phi}(\theta)^* / g_{x0}(\theta)g_{x0}(\theta)^*$ for a single element using $\phi = 0$, which is a small value between 10^{-2} and 10^{-3} . For synthesis imaging we need the ratios $Q(\theta) / I(\theta)$, $U(\theta) / I(\theta)$ and $V(\theta) / I(\theta)$ using cross-correlation between two pairs of orthogonal elements. We find for the relative linear polarization values with order of magnitude given by $|E_{-\phi}(\theta) / E_{0\theta}(\theta)| \sim |g_{x\phi}(\theta) / g_{x0}(\theta)|$ which increase quadratic with θ (as will be shown in the next subsection) and reach a value 0.5 at half power by comparing figure 3.9 with figure 3.8.

3.6.3 Polarization of a phased array station beam

The station beam of a LOFAR phased array station has a full width at half maximum of at most $\sim 13^\circ$ and cuts out only a small section of the polarized structure of the element antenna pattern. The side lobes of the station beam cut out a different section of the element beam and get accordingly a different polarization.

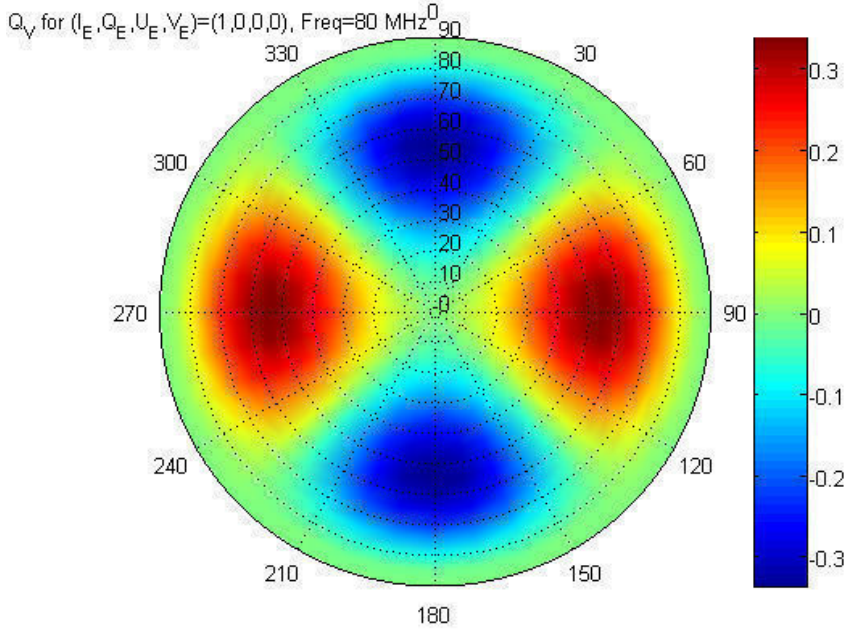


Figure 3.9. Polarization pattern in Q (relative to peak) total intensity with 2-fold symmetry of a dual polarized LOFAR LBA antenna for an un-polarized input wave at 80 MHz as function of azimuth and zenith angle (From [Arts, 2005])

A station beam that tracks a sky source follows a trace in azimuth and elevation over the polarized element antenna pattern. So the polarization characteristic of the phased array station beam that tracks with less than 0.25° /min changes continuously, but shows only little change with time since the element beam changes over much larger angular scales as shown in figure 3.8.

When a source field is tracked during a short synthesis observation we have to deal with different effects. These effects could be visualized by projecting figure 3.6 with different short synthesis fields along the track of a long synthesis over the polarization pattern in figure 3.8 that covers a hemisphere. One effect relates to the rotation of the source field relative to the coordinate system of the snapshot image. The other effect is that the polarization angle is relative to Zenith instead of the Celestial pole, which means that the polarization angle of each object in the field has to be

rotated over the parallactic angle. The continuous rotation of the U'_0, V'_0 -coordinates as discussed in subsection 3.5.4 eliminates the rotation of the image in the field during the tracking interval. This rotation correction is exact at the centre of the field but differential position effects over the field can be ignored for synthesized snapshot images shorter than ~ 10 min as discussed in 3.5.4.

Correction of the polarization angle requires a separate correction of the coherency matrix that contains the four observed polarization visibilities of each baseline sample [Hamaker, 1996]. For a field centred at the pole we need a correction for the polarization angle at each image pixel to realign the polarization angle for a coordinate system centred at the pole instead of Zenith as is the case for the antenna signals. In this special case the rotation for each pixel as function of time is the same and we need a single polarization rotation correction for all data of each synthesized snapshot image.

For snapshot images at lower declination the polarization rotation is different for each pixel but the change in differences during ~ 10 min can be ignored. In addition, as discussed in subsection 3.5.4, the polarization rotation during ~ 10 min is small such that degradation of polarized intensity is less than 0.008 % and can be ignored as well. So, a correction per pixel per synthesized snapshot image is therefore only required per synthesized snapshot image before these are combined to a single synthesis image. These corrections have indeed to be made in the image domain and we need four images, one for each observed polarized coherence. A single Mueller matrix per pixel could apply the required rotation correction including conversion to the four Stokes parameters as well as correction for beam polarization. The latter is true, since the polarization characteristics are sufficiently identical for all stations, which have almost identical element antenna orientation.

Faraday rotation by the ionosphere is not identical for all stations, which means that the polarized visibilities need a separate rotation correction for each visibility, and even per source direction. The differences in Faraday rotation are proportional to differences in phase as caused by refraction but also proportional to wavelength and will be discussed in section 4.1. A TEC difference of 0.1 TECU gives 24 rad phase difference at 35 MHz and ~ 1 rad differential Faraday rotation, and at 70 MHz just 12 rad phase but only ~ 0.25 rad differential Faraday rotation. Such TEC differences occur over the FoV of a LOFAR station beam but also between stations with separations larger than 10 km and could be caused by larger scale structures in the ionosphere. However, TIDs could cause differential variation of 0.05 TECU in 10 min but tracking of a field at 45° elevation could cause a change of 0.08 TECU/min along the line of sight. LOFAR needs corrections for Faraday rotation by the ionosphere that are not only different per station but also different per image pixel per synthesized snapshot.

The fast polarization rotation common to all stations and to the whole field needs a correction at least every min and could be applied per visibility. The slower change that varies with position needs a correction once per ~ 10 min and could be com-

binning with parallactic and beam corrections discussed above and applied per pixel per synthesized snapshot image.

Although a complete polarization correction scheme is outside the scope of this discussion, we give an order of magnitude estimate of the instrumental polarization over the FoV of a synthesis image.

To this end we model the power response pattern of the average dipole element in X-direction to an un-polarized signal by an elliptical profile in azimuth, which can for frequencies below 50 MHz be approximated by

$$V_{XX} = \cos \theta (\cos^2 \phi + \sin^2 \phi \cos \theta) \quad (3.68)$$

where ϕ is the azimuth angle and θ the zenith angle. For the orthogonal Y-element we get

$$V_{YY} = \cos \theta (\sin^2 \phi + \cos^2 \phi \cos \theta) \quad (3.68a)$$

The total intensity given by observed Stokes parameter I_V equals

$$I_V = V_{XX} + V_{YY} = \cos \theta (1 + \cos \theta) \quad (3.69)$$

which is indeed independent of azimuth angle.

The polarization given by observed Stokes parameter Q_R equals

$$Q_V = V_{XX} - V_{YY} = \cos 2\phi \cos \theta (1 - \cos \theta) \quad (3.70)$$

The difference in the shape of the element power beam for XX and YY coherencies creates after subtraction of XX and YY images an observed relative polarization over the field of the station beam given by

$$Q_V / I_V = \cos 2\phi (1 - \cos \theta) / (1 + \cos \theta) \quad (3.71)$$

Near the Zenith we can approximate the polarization of the element beam (3.71) by

$$Q_V / I_V = \frac{1}{4} \theta^2 \cos 2\phi \quad \text{for } \theta \ll 1 \quad (3.72)$$

For U_V / I_V a comparable relation is found for the LOFAR antennas using $\sin 2\phi$ instead of $\cos 2\phi$.

This is the same quadratic property as for most dish telescopes that have small beam polarization close to the centre of the main beam (with θ expressed in fractional width of the station beam).

Equation (3.72) assumes that the XX and YY channels are properly calibrated for the centre of a dish telescope or for the Zenith direction of a phased array station

respectively and suggests that the observed polarization over the phased array station beam could strongly increase when pointed at larger zenith angles, but can be removed by appropriate calibration.

3.6.4 Polarization over the phased array station beam after gain calibration

We consider a station array that is properly calibrated for the centre of its station beam when that is pointing at Zenith where the element beam pattern has no polarization. When the station beam is subsequently pointed towards an un-polarized calibration source at θ_0 and $\phi_0 = 0$ we get responses according to (3.68) and (3.68a) that require additional calibration factors $(1+a)$ and $(1-a)$ such that the calibrated responses V'_{XX} and V'_{YY} for this source are equal for both observed coherences. This implies $V'_{XX} = (1+a) V_{XX}$ and $V'_{YY} = (1-a) V_{YY}$ for the whole beam which gives an un-polarized response for an un-polarized source at the centre of the beam. For a source at θ and $\phi = \phi_0 = 0$ we find however polarization Q'_V and intensity I'_V and instead of (3.72) we get

$$Q'_V / I'_V = (\cos \theta_0 - \cos \theta) / (\cos \theta_0 + \cos \theta) \quad (3.73)$$

Inserting $\theta = \theta_0 + \delta\theta$ we get after linearization

$$Q'_V / I'_V = \frac{1}{2} \delta\theta (\frac{1}{2} \delta\theta + \tan \theta_0) \quad \text{for } \delta\theta \ll 1 \quad (3.74)$$

If the centre of the station beam is pointed at Zenith with $\theta_0 = 0$ while $\phi_0 = 0$ we find back our result (3.72) with quadratic increase with $\delta\theta$. If the station beam is pointed at lower elevation θ_0 while $\phi_0 = 0$ and then recalibrated using an un-polarized source at that location, we find a different polarization pattern over the station beam. The relative polarization in Q'_V then increases in proportion to zenith distance $\delta\theta$ from the beam centre and is proportional to the tangent of the zenith angle (ignoring the $\delta\theta$ term between parentheses in (3.74)). This is an interesting result that shows how independent calibration of XX and YY channels on an un-polarized source could lead to proper polarization calibration for I and for Q. Since we increased the X-gain and decreased the Y-gain by small and equal amounts, the XY and YX gains will hardly be influenced and the -gain- calibrated U'_R and V'_R will hardly differ from observed U_V and V_V respectively. In fact we need (3.67) and (3.67a) and combine not only XY and YX channels but all four observed polarization coherences to obtain the four true Stokes parameters. The analysis of Q_V for an un-polarized source used in fact only the first two elements of the first row of a full Mueller matrix.

Since the shape of U_V as response to an un-polarized source resembles the Q_V -pattern [Arts, 2005] rotated over 45° we can expect a similar result as (3.74) if full polarization correction is obtained for a single position in the field of the station beam. This correction is also a good approximation for other points with $\delta\theta < 0.1$

within the station beam. More specific, when all points in a facet beam are corrected with the Mueller matrix for the centre of that facet beam, we expect only small residual beam polarization $(Q^2 + U^2)^{1/2}$ for which the order of magnitude is given by (3.74). Further analysis is needed to show the effect of a shift $\delta\phi$ that involves the $\cos 2\phi$ factor for Q in (3.72) and the $\sin 2\phi$ factor for U and how this needs to be combined with $\delta\theta$ to define a residual polarization as function of the radial distance to the centre of the facet. Since the residual effects are small, linearization could provide in principle an efficient correction procedure with sufficient accuracy [Brouw, private communication].

The polarization of the average element beam pattern provides a contribution to the intrinsic measured polarization of sources in the station beam which could be removed in two steps. First we need proper correction of the station beam shape for each image made in each of the four polarization components of the observed visibilities. This beam shape is the product of the scalar station array pattern and the I pattern of the element antennas over the area of the station beam and is used for each of the four coherency images. In a second step proper corrections for polarization rotation and polarization conversion need to be made based on a description of the element pattern where the elements of the Mueller matrix are normalized for the I contribution. The observed polarization coherence-vector components for each pixel in the synthesized snapshot images are converted to four Stokes parameters in a coordinate system for the final synthesis image. This correction in the image domain assumes that all stations are almost equal. In case that station beam patterns are not equal, a quasi-convolution could be applied that gives the nominal amplitude pattern of a facet beam a distortion that corrects for the amplitude variation in the station beam over the extent of the facet beam.

We need to realize that such a station beam correction in the image plane also corrects synthesis side lobes at their apparent location in the image, while their actual polarization is determined by the polarization at the location of the object that emanates the side lobe responses. The disturbing effect of these side lobes can only be reduced by creating low side lobes either by proper tapering of uniformly distributed U, V -samples or, preferably, by subtracting the source from the visibility data before direct imaging.

Simple relative calibration of the XX and YY visibilities using a single un-polarized source near the centre of a facet field already provides much lower beam polarization in Q for a synthesis array with almost identical phased array stations than for an array with conventional dish telescopes. The relative instrumental Q polarization over the FoV defined by the station beam then increases linearly with distance $\delta\theta$ from the reference position near the centre and has a small slope that depends on the zenith angle θ_0 of the beam. According to (3.74) the relative Q polarization of the compact LBA station beam at 50 MHz is about 10 % at the quarter power level when it is pointed at 45° zenith angle. For the HBA stations with their factor 3 nar-

lower station main beam, polarization at quarter power level will be a factor 3 lower. Indeed LOFAR synthesis images showed “unexpectedly low” instrumental Q polarization effects especially for observations near Zenith [A.G. de Bruyn, private communication] with a magnitude indicated by (3.74)

This result for phased array stations should be compared with dish telescopes that have typical 50% relative polarization at the quarter power level of the beam.

Since self-calibration on an un-polarized source gives already first order correction for beam polarization that converts I into Q, the remaining corrections are small and their variation over the field of a station beam is even smaller as indicated by (3.73) and figure 3.8. A more careful analysis of the azimuth dependence shows that (3.73) is indeed the dominant term and that higher order terms are much smaller [Hamaker, private communication].

When a full polarization correction for a specific position of the element beam pattern is performed, this will just as for the situation analyses for Q be approximately correct for its nearby points covered by the station array pattern. We therefore conclude that there is no need for a more frequent update of full polarization correction faster than once per 10 min, which is adequate to allow only very little degradation in polarization intensity due to rotation as discussed in section 3.5.4, while the rotation angle is not influenced. In subsection 3.6.3 we discussed the effects of Faraday rotation and concluded that tracking at an elevation of 45° at a frequency of 35 MHz requires a rotation correction per antenna station that should be applied in the visibility domain once a minute. This results in a fast rotation correction for the whole image field, while the slower varying distortions over the field need only a correction once every ~ 10 min.

An important aspect is that although proper polarization can be obtained with full matrix correction procedures, signal to noise ratio is lost in the final answers if one of the observed components has a high weight but a low signal-to-noise ratio contribution to a specific Stokes component. The important result is that antenna beam polarization itself does not limit polarization purity in a calibrated and corrected image but only effective polarization sensitivity. Such degradation starts to play a role for phased array antenna station for observations below 30° elevation where the sensitivity of orthogonal dipole-like antennas is at least reduced by a factor two compared with zenith direction and where their sensitivity ratio could exceed a factor two.

3.6.5 Element beam pattern and blind angle effects

Electro Magnetic (EM) coupling between antennas in an array causes two effects. The first is that the beam pattern of each individually excited antenna, while all other elements are not excited, differs from the beam pattern of a free standing antenna by up to 30 % [Cappellen, 2006]. The second effect is the so-called mutual coupling where current in one element induces voltages in all other elements that are connected by the so called mutual impedances to these elements. The beam pattern of an array can for each direction be evaluated by vector summation of the field contribution of each antenna beam in that direction as follows from each individual excitation.

An incident plane wave induces voltages with amplitudes given by the individual beam pattern of each antenna and a phase that depends on the direction of the wave. Mutual coupling between elements creates an additional voltage in each element induced by the currents in surrounding elements as defined by mutual impedances, self-impedance and load impedances of the antennas in an array. An element at distance R from a reference element contributes a coupling signal proportional to the current induced by external signals and is inversely proportional to its distance from the reference element. Since the number of contributing elements in a regular array increases not only proportionally with R , but could for specific directions and specific frequencies have a constant phase difference, grating like phenomena will occur, the so called blind angles. The additional signal on each element has a fixed relation in amplitude and phase with it surrounding elements and could for a large array be described as a convolution in the spatial domain. Such a convolution provides an additional beam that is multiplied with the array beam just as the average element beam pattern. In a small array however, every antenna has a different environment, which means that convolution is only a first order description.

An array with N elements has N element antenna patterns and an N^2 impedance matrix. A tedious EM simulation is required to determine all the patterns and the impedance matrix of an array for a large set of frequencies. A simplified method has been developed allowing first order estimation of the blind angle effect [Wijnholds, 2008] and some results are repeated in figure 3.10. The simulation used a uniform element beam, which means that the average beam pattern that resembles the freestanding element patterns in section 3.6.2 has to be multiplied with the blind angle pattern.

For an array with randomized element positions such as the LBA we get an average element beam pattern with little fine structure, however when the same elements are placed in a regular array we find that all element beams have almost equal fine structure with deviations up to 30% from the average value. Moreover, these fine structures are only slightly wider than a station beam [Cappellen, 2006].

The station beam of a regular array such as the HBA could therefore appreciably change when it tracks the sky. A beam of 3.5° needs $\frac{1}{4}$ hour to traverse and could suffer up to a 50% change in sensitivity. The pictures in figure 3.9 indicate however that such large changes occur at a specific frequency only in two places that have an extent equal to the station main beam. At other frequencies the changes are smaller but occur over a larger area. A simple graphical integration of the total “solid angle” in l,m -space weighed with its depth provides values of 1, 1.6 and 2 times the main beam area at 150, 180 and 200 MHz respectively. Interestingly these numbers are the same as for the grating lobes of an array that has an isotropic element pattern, which is also assumed in the simulations for the blind angle.

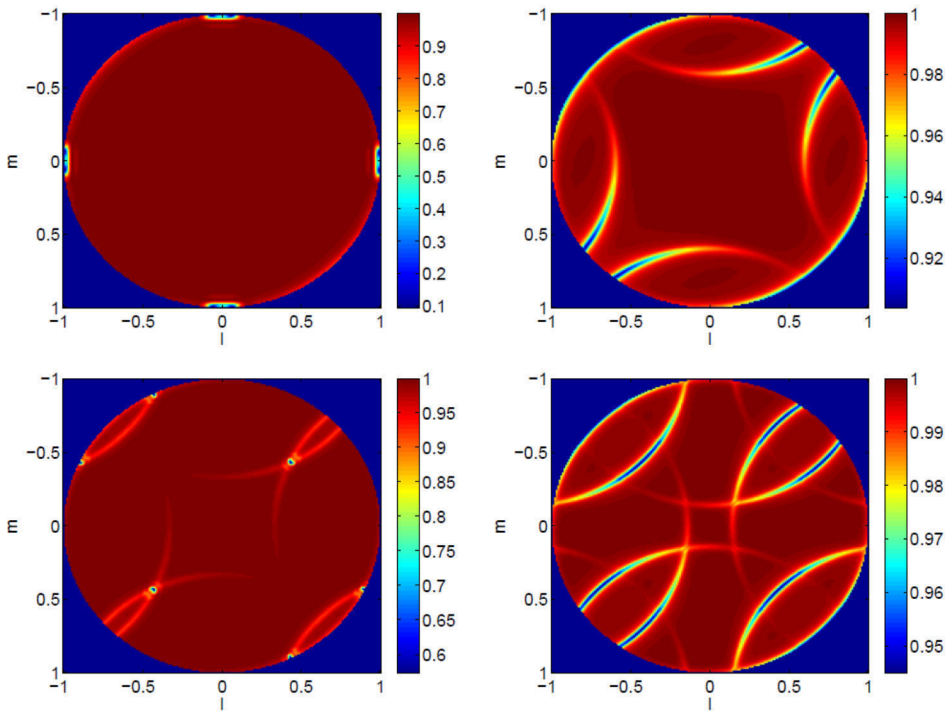


Figure 3.10. Array gain over the sky for a 96-tile array of x-dipoles without inter-tile spacing at 120, 150, 180 and 210 MHz respectively assuming a MIMO coupling model [Wijnholds, 2008].

Since the width of the blind angle structure is comparable to the width of the station beam, serious distortion could be expected that can no longer be modelled with a standard model for the main beam. On the other hand the beam pattern of an array is mainly determined by the spatial distribution of its element antennas and to a lesser extent to the effective weight of the antenna signals that are disturbed by the

mutual coupling. As a consequence the top half of the station beam is still described by a symmetric Bessel function that could decay into an asymmetric side lobe pattern. We conclude therefore that the blind angle phenomenon could cause changes in the shape of the station main beam when a sky field is tracked. These changes could in principle be modelled with a simple amplitude factor for the centre part of the station main beam that will, however, be handled by self-calibration. The outer part of the station main beam below half power will also vary in shape, which could be corrected in principle using a spatial convolution correction for each baseline with an affected station.

The effect on synthesis imaging is reduced by rotating the configuration of each station. However, when a residual grating lobe disturbs the self-calibration of a particular short synthesis image, it would be better to delete the baselines that are affected by the station that is the prime cause. This approach could also be used for stations that pass a blind angle that would seriously deform a station beam such that appropriate modelling and correction is not possible with simple functions.

In a synthesis image with order 40 stations a 4% dip in one of the station beams causes at most a 0.1% dip in the outer half of the average station beam pattern below half power of a short synthesis observation of about 10 min. This figure is low when compared to a change in beam shape of more than 1% by changing foreshortening during tracking over 10 min and will be further reduced in a long synthesis observation. We finally conclude that blind angles due to mutual coupling effects do not need correction. It could be handled by deleting visibility data of facets that show a deviating station gain factor for only a part of the station beam.

3.6.6 Combining stations with different polarization characteristics

In the previous section we mentioned that rotating a regular station configuration for each station could reduce the average effect of blind angles. Indeed LOFAR station configurations are rotated not only to reduce possible blind angle effects but also to average the effect of grating lobes on a synthesis image [Wijnholds, 2008]. If the whole station would be rotated, then also the orientation of the element antennas between stations would change, and indeed the beam formers in the LOFAR stations have the option to recombine the signals of X- and Y-antennas such that a specific polarization in the direction of the station beam could be obtained. This option could even support different orientations of antenna elements in a phased array station. Indeed, such a polarization diversity scheme has been contemplated for the Low-band stations and would provide an averaged element beam pattern that is rotation symmetric for station signals. Such a pattern could be attractive in principle when phased array beams would be made by combining stations in the core area of LOFAR. However, at the time that a decision on the station layout had to be taken the consequences for control and correction procedures could not be estimated by lack of sufficient evidence of the impact of polarization diversity

schemes. As a consequence the choice was made to rotate the station configuration, but to counter rotate the element antennas such that all stations would have the same polarization pattern on the sky. This was considered important since all baselines would then see, apart from differential Faraday rotation between stations, the same polarization for an object and would need the same polarization correction that would vary only gradually for different directions on the sky. This means that polarization corrections need not be made per baseline in the visibility domain but could be made in the image domain after Fourier inversion.

The rotation of the stations is organized in such a way [Bregman, 2012], that each range of baselines that is provided by a combination of rotated station configurations has a reasonable distribution of grating lobes. A simple method has been used based on the configuration of the synthesis array where stations are grouped in rings centred on the central cluster of six stations that forms the so called super-station. The six stations in each ring have a uniformly rotated side lobe pattern that is interspersed with the grating lobes of subsequent rings and reduces the grating lobe pattern of the station beam averaged over the stations in a ring.

An important aspect of this rotation scheme per ring is that every range of baselines that optimizes the brightness sensitivity for a specific sky field also has a properly averaged station beam.

3.6.7 Combining beams of stations with different diameter

In principle problems with beams from stations with different aperture sizes could be avoided by appropriate tapering of all stations during an observation to make them effectively equal and even circular [Hamaker, private communication].

This leads to sensitivity loss for the largest stations and complicates calibration especially on the longest baselines where these stations appear. More practical is an approach that uses spatial filtering after observing by using convolution of U,V-data just as for creating small facets. This approach considers unequal beams as a problem that needs to be handled by appropriate imaging algorithms and not by mutilation of a station.

As already said in the introduction of section 3.6 Fourier imaging with different station beams effectively leads to varying effective taper coefficients for the visibilities that form a synthesized beam of each source depending on its location in the FoV. Fortunately the polarization response for each station is in principle the same since it is caused by the element antennas. Station rotation and counter rotation of the elements could however cause minor differences. In practice there will be differences between stations at different geographical location, since their local Zenith points at a different location in the sky. For stations at 600 km distance from the centre of the array this is about 5° .

When facet imaging is used we have to deal with a different gain slope over each facet as determined by the relevant part of the station beam. A simple additional convolution filter that corrects the amplitude of the affected visibilities could in principle correct this. Only a first order correction is needed that limits the differences in effective taper coefficients for a point source as function of its distance from the facet centre.

The central part of the station main beam above half power can be accurately described by an elliptical Bessel function. This shape is determined by the largest separation between elements in a station aperture and is hardly affected if intermediate element antennas in a station fail. Even more important for processing efficient imaging is the fact that the polarization pattern over the sky is almost identical for all stations since the antenna elements in every station are similarly oriented. The main beam of each station and also the station side lobes view the sky through a polarizing pattern that has only large scale variation as determined by the average element pattern. When the scalar station beam tracks the sky then also the scalar side lobes get the polarization as determined by the element pattern. Only when the array patterns of X and Y are different, for instance by improper station calibration or by element failure, polarization will be observed since X and Y channels have large different gain for a source at a location with different X and Y side lobe pattern. This is contrary to the side lobe pattern of a dish, where the dish transforms the illumination pattern giving every side lobe a polarization structure just as the main beam. A tracking dish telescope gives a rapidly varying polarized response when its polarized side lobes move over un-polarized sources outside the main beam. A tracking phased array sees only slowly changing polarization by un-polarized sources that move by Earth rotation through the much wider polarization structure of the element pattern.

Another difference is that dish telescopes have only two receivers that amplify the two polarized antenna signals, while a phased array station has two receiver sets. Individual receiver gain changes of the HBA tiles caused by switching of delay lines average out quickly if the switching is not done identically for all tiles simultaneously. In practice, the stability of the complex gain of the antenna receiver chain has a much longer time scale than the ionosphere and both will be properly handled by the self-calibration approach.

3.6.8 Summary and Conclusions

The results of previous sections can be summarized as follows:

- A synthesis image needs beam shape correction, polarization correction as well as parallactic polarization rotation correction and Faraday rotation correction, as function of position that could be constant during a 10 min synthesis period. Such corrections need all four coherence images (XX, XY, YX and YY) to form proper images per Stokes parameter (I, Q, U and

- V), and could be applied in the image domain of a synthesized snapshot that should have comparable duration.
- Average Faraday rotation by tracking of the station main beam is different for each station and varies rapidly when a field is tracked at elevations of $\sim 45^\circ$. This rotation could be assumed constant over the station beam and needs at 35 MHz correction of all 4 visibilities at least every min.
 - Polarization is determined by the element beam, which is the same for all interferometers independent of station size. Since the elements in all stations have almost the same orientation, simple polarization beam correction is possible after averaging all interferometers by an imaging process (even for the side lobes and even for different station sizes).
 - Differences between the station voltage main beams g_{ik} as caused by rotation of the element configuration are small for stations of equal size. So it is justified to assume just a single power pattern $g_k g_k^*$ that is some average over all stations.
 - The main effect of combining stations of different size is that visibilities of point sources in a synthesis array vary with the location of the source in the field. This effect could in principle be reduced by introducing a complex gridding convolution that corrects station baselines not only for non-planarity but also for differences in amplitude variation over each facet beam.
 - This approach is a sound basis for hybrid imaging methods, where the differences between individual station beams are properly taken into account when strong point sources are subtracted from the U,V,W-visibility dataset and where the residual visibilities are imaged using regridding with convolution corrections and a 2-D FFT.
 - Bi-scalar (separately on XX and YY channels) self-calibration on a single un-polarized source near the centre of the station beam provides zero Q at that location and shows relative beam polarization in Q that increases linearly toward the edges of the field. Less than 1% instrumental Q is expected at quart power level in the beam of the small HBA stations when no further beam polarization correction is made per snapshot.
 - However U and V are not corrected by such a bi-scalar approach. This means that a renormalization is required that accounts also for the phase difference between X and Y-channel of each telescope before a nominal Mueller matrix for a specific direction in the element beam pattern can be applied that corrects all Stokes parameters.
 - This renormalization includes the nominal beam polarization at the location of a calibration source and could even include the polarization of that source leading to full polarized calibration.
 - With proper polarization correction for a single position within the station beam the polarization distortion increases only linearly with distance from this position.

- Antenna beam polarization itself does not limit polarization purity in a calibrated and corrected image but reduces only effective polarization sensitivity (3.6.4).
- Receiver gain difference between the two polarization channels of a station beam is the average of a large number of element receivers varies only slowly over time and can be properly self-calibrated.
- In a synthesis image with order 40 stations a 4% blind angle dip in one of the station beams causes only a 0.1% dip in the lower half of the average station beam pattern of a short synthesis observation of about 10 min, and could be ignored. The more important top half is however properly corrected by self-calibration.
- The most effective mitigation approach for blind angles is just deleting baselines that are potentially affected by a blind when a sky field is tracked.
- The same holds for disturbing residual grating lobes that could distort self-calibration and imaging of individual snapshot images.

The most important conclusions are:

- A phase array stations has much smaller polarization variation over its station beam than a dish antenna.
- After polarization calibration of a phased array station beam for one or more positions, the polarization errors grow approximately linearly with distance to these reference positions.

3.7 Comparing processing for 3-D, 2-D and Synthesized snapshot imaging

In this section, various imaging approaches are compared. It will be shown in chapter 4 that the LOFAR stations have sufficient sensitivity to observe a number of sources per station beam that allow proper self-calibration for the whole beam. In chapter 5 we will analyse how many sources have to be subtracted accurately, using this calibration, to reach the thermal noise in a wide-band continuum image. These numbers vary from 5 to more than 1000 depending on the array configuration that determines the average side lobe level in a synthesis observation. In the latter case the processing for image forming is fully dominated by the subtraction process, and we will investigate in this section what the actual balance is between the various processing steps in the image forming process.

In the first place it will be shown that convolution processing (for continuum imaging) dominates over Fourier inversion, even for the proposed synthesis imaging

approach using ~ 40 synthesised snapshot images in the reference plane of the array. The latter approach uses complex convolution correction and does not require faceting when stations are closer than 80 km from the core, as is the case for Dutch LOFAR configuration. Conventional polyhedron imaging at 50 MHz using a small convolution kernel to allow a small FFT for each facet would require already ~ 850 facets and leaves second order phase errors up to ~ 0.3 rad, on the longest baseline for sources at the edge of the field. Using a small but complex convolution kernel that corrects for these second order terms requires ~ 470 facets and makes conventional imaging with W-axis towards the centre of each facet a practical alternative. In contrast, guessed estimates for conventional image forming based on 3-D imaging or 2-D imaging with W-projection to obtain a FoV that covers the station beam at 50 MHz with a single facet would even exceed the processing capacity for correlation using stations out to 3 km from the centre of the array.

In previous sections, we have introduced the Complex Multiply Add (CMA) operation as a metric for processing volume required by a program to execute large sets of operations. Typically 6 floating point operations (flop) are required to execute a single CMA. Processing power of a platform is expressed in flop/s and is only one of the processing requirements for a platform that has also to provide adequate data throughput rate and intermediate storage to perform certain tasks efficiently such as cross-correlation, convolution, fringe tracking, Fourier transformation, and interpolation on large data streams.

Apart from executing a small processing kernel on a large dataset there are additional operations to determine the coefficients of the kernel. For synthesis imaging with N_{st} stations that have 2 polarization channels we form $2N_{\text{st}}^2$ complex visibilities for N_{ch} spectral channels. Current imaging packages were developed for $N_{\text{st}} < 30$ and $N_{\text{ch}} < 10^3$ at typical read out periods of 10 s and processing algorithms and program code has been optimized for producing image fields with $N_{\text{p}} < 10^7$ pixels on a single PC type platform. LOFAR has $40 < N_{\text{st}} < 80$ with $N_{\text{ch}} \sim 10^5$ and 1 s read out, providing 10^4 times more visibilities per unit time. Typically 10 times more continuum images are formed to cover the extended bandwidth that are a factor 10^2 larger, requiring the equivalent of a cluster facility with 10^4 high performance laptops to keep up with the output data rate of the correlation platform. SKA will even have 3 – 30 times more stations providing 10 - 10^3 more baselines and full beam images that have at the higher frequencies a factor 10^2 more pixels. This requires not only proportionally increased processing power but also an increased data throughput by another factor $> 10^2$.

The organization of the correlated data in visibility streams per facet beam and per spectral channels allows a high degree of parallelization and seems straightforward. However, routing of the massive data streams from stations to correlation processes and from there to imaging processes asks for optimized platform architectures. Not only the structure of processing platforms has to be optimized, also the structure of the programs that have to deal with a different balance between kernel oper-

ations per visibility and kernel operations per image pixel will change. Even more important, total processing power is no longer determined by operations that provide calibration parameters per station as is the case in most legacy packages, but will be dominated by correction of the visibility stream. Indeed performance tests with newly developed processing software for LOFAR shows that laptop platforms are not limited by throughput and memory requirements but limited by kernel CMA requirements. This shows that simply estimating the CMA capacity to complete a task such as Fourier transformation is enough to estimate the equivalent number of processing units in a HPC platform that are required to complete this operation in a given time.

In this section we will compare different methods that make synthesis images with a wide FoV that are potentially suitable for LOFAR and SKA. We strike a balance between the various operations that are needed in visibility domain and in image domain that minimizes total CMA requirements just based on CMA requirements of each type of operation.

In previous sections we have seen that aperture synthesis imaging can be realized by Fourier transformation (FT) of the observed visibilities formed by correlation between antenna pairs. In a generalized approach [chapter 19, Taylor, 1999] we need a 3-D FT that transforms the 3-D baseline set of an Earth rotation synthesis observation from which a 2-D image can be obtained. We need a series of 2-D transforms that provide quasi-images and an interpolation along the n -axis is needed to form the final image on the spherical l, m, n -surface with a limited FoV as determined by the extent along the n -axis. Planar arrays need only a 2-D FT and a single planar l, m -image is obtained with a FoV that can cover one hemisphere unambiguously. A long synthesis observation could then be made as a sequence of short ones and could then cover even more than a hemisphere. Practical arrays that are planar to first order could still use a 2-D FT, but the accuracy of the images is then confined to a smaller solid angle on the sky that could even be smaller than the extent of the beam of an antenna station. In that case the field of the station beam observed by the visibility function could be further reduced by a convolution operation, such that only a facet field remains that can be imaged by a 2-D FT with limited extent that is almost distortion free. When a complex convolution kernel is used even second order corrections for the non-planarity of the baselines between the stations can be obtained that extend the size of such a facet field.

Efficient imaging can now be realized by defining an appropriate set of facets that can be imaged using a 2-D Fast Fourier Transformation (FFT). Such a FFT requires that its input is defined on a rectangular grid, which needs a regridding convolution operation of the observed baseline visibilities. Fortunately such a convolution can be replaced by a complex one that corrects for 2nd order terms and extends the facet size for which the 2-D FFT will provide an image of which the accuracy is limited by higher order terms. The processing needs for imaging has therefore two components, one for convolution of the observed data and regridding these on a rectangular grid and one for the FFT. Processing efficient imaging needs therefore

to strike a proper balance between the two operations, while additional operations for combining the facets are smaller and will be ignored.

Most conventional imaging packages have chosen a coordinate system with its W-axis towards the field of interest, which transforms an Earth bound planar 2-D array into a 3-D space array when a sky field is tracked during a synthesis observation. Conventional polyhedron imaging defines for a full synthesis observation a large set of small facet images that each need a single small 2-D Fourier transform with a W-axis towards the centre of the facet field. The size of the facets is determined by external non-planarity of the baselines that emerged as the consequence of the tilt of the array plane relative to the chosen reference plane for the field that is tracked. The most recent approach called W-projection uses only a single facet field that covers the station beam but needs a large complex convolution kernel of which the linear extent is proportional to the maximum baseline.

Unfortunately, the processing power (in flop/s) required by these imaging packages to complete a synthesis image in a time comparable with the observing time becomes too large for LOFAR because of its large FoV, high resolution and large number of baselines and large number of spectral channels that need to be used in a continuum image. The main reason is that no use is made of the intrinsic planarity of the 2-D array, since the focus of the conventional packages has been on dealing with extrinsic non-planarity in an attempt to work with a single FFT for a full synthesis observation where a planar array tilts during tracking of a sky source. Such an approach is indeed justified for line imaging where a large number of Fourier images has to be made each with only few visibilities as input.

Our synthesized snapshot approach uses a coordinate system with its W-axis towards Zenith of the centre of the array and needs only a small complex convolution kernel that corrects for intrinsic non-planarity caused by Earth curvature. However, the maximum tracking time for a synthesis image in such a coordinate system is by Earth curvature limited to about 10 min, for arrays with stations out to 80 km from the core. The synthesized snapshot approach with its inherent rescaling and rotation of each image before integration to a final long synthesis fortunately allows correction for average beam effects over that short interval, which is especially important for arrays with phased array antenna stations. These corrections have different components of which a part could indeed most easily be implemented in the image domain. The required number of synthesized 2-D FFT snapshot images is inversely proportional to their duration that is limited by beam size and non-planarity. However, field and polarization rotation need a full image correction only once per 10 min, which is matched to the maximum duration of a synthesized snapshot image with the Dutch LOFAR configuration covering its largest beam. Interestingly, the necessary convolution operation defines a kernel diameter (3.48) that is proportional to the square of array extend and necessitates faceting for a larger array configuration. Although a combination of snapshots and facets is possible in principle it complicates the imaging process but might be attractive for arrays with

stations further than 80 km from the core by requiring much less facets than polyhedron imaging, even when enhanced with complex convolution correction.

In view of the different scaling laws for 3-D Fourier inversion using a real convolution kernel and 2-D facet imaging and 2-D snapshot imaging using a complex kernel, a more detailed analysis is needed to strike a proper balance for minimum processing requirements.

In subsection 3.7.1 we give a more detailed analysis of the different contributions to the total required processing capacity for each of the three approaches by convolution, transformation and interpolation as function of FoV, resolution, number of baselines, number of spectral samples per baseline and number of temporal samples per baseline.

In subsection 3.7.2 we compare the imaging methods for different applications with focus on LOFAR. In subsection 3.7.3, we compare the required processing power for real time imaging with the processing power for cross-correlation to allow a first order estimate for the magnitude of the platform for post correlation processing compared to the magnitude of the correlation platform.

We summarize conclusions in subsection 3.7.4.

3.7.1 Processing capacity of the main steps in hybrid imaging

In the following subheadings we introduce the basic elements that together define the total processing capacity required for creating a synthesis image in Complex Multiply Add (CMA) operations. Our purpose is to compare the processing required for straightforward 3-D imaging with that for two types of 2-D Facet Imaging. The first type is enhanced polyhedron imaging where the extrinsic non-planarity caused by projection of baselines is corrected by a complex convolution. The second type uses fewer facets but each facet needs a set of 2-D FFT images called synthesized snapshots that each need only convolution correction for the intrinsic non-planarity caused by Earth curvature. In cases with a narrow station beam only a single facet could suffice, and for very short baselines even extrinsic non-planarity could be covered by a complex convolution kernel of limited size.

This comparison assumes that all necessary beam shape and polarization corrections that need to be done on time scales shorter than 10 min can indeed be implemented in the visibility domain with complex gain corrections and convolution per baseline and can indeed be handled by a kernel size of 7^2 pixels only. This constraint on the size of the processing kernel determines the number a facets, which itself does not seriously increase the total processing load but complicates the structure of an imaging package and it's partitioning over the processing nodes of a HPC platform.

In the following subsections, we address the various processing aspects and their contributions to the processing load.

3.7.1.1 Resolution and FoV determine number of visibility samples

Resolution and FoV determine the required sampling of a correlation interferometer and have been analysed in section 3.2. Objects away from the centre of a station beam suffer from amplitude degradation in an interferometer that tracks a source in a rotating sky. For stations with diameter D that use parabolic amplitude taper and for baselines with length B is a maximum amplitude degradation of 1.7 % is tolerated for sources at half power with integration time τ and bandwidth $\delta\nu$ ((3.30) and (3.32) respectively) we found

$$\begin{aligned}\tau &< 2323 D/B & [\text{s}] \\ \delta\nu &= 0.168 \nu D/B\end{aligned}$$

For a continuum image with total bandwidth $\Delta\nu$ [MHz] we get for baseline B in synthesis time T_s [s] a number of time samples N_t each with a number of spectral channels N_c given by

$$\begin{aligned}N_c &= \Delta\nu / \delta\nu \\ N_t &= T_s / \tau\end{aligned}$$

For baselines shorter than B we could in principle work with longer integration times and wider channel bandwidth to reduce the output data rate of the correlation process. This would however introduce image distortions that are avoided if such integration is done by the gridding convolution. In practice we work however with constant values as determined by the longest baseline B_{\max} and find for a continuum image a total number of samples N_{sa}^c provided by N_b baselines given by

$$N_{\text{sa}}^c = N_b N_t N_c = N_b (\Delta\nu / \nu) (B_{\max} / D)^2 T_s / 390 \quad (\delta\nu < \Delta\nu) \quad (3.75)$$

This equation assumes that the frequency coverage $\Delta\nu$ for an image that is covered by a number of narrower channels with bandwidth $\delta\nu$, which is required to avoid bandwidth smearing. For line imaging we could have the situation $\Delta\nu < \delta\nu$ and we get $N_c^L = 1$ and

$$N_{\text{sa}}^L = N_b N_t N_c^L = N_b (B_{\max} / D) T_s / 2323 \quad (\text{for } \delta\nu > \Delta\nu) \quad (3.75a)$$

For a typical line observation we have $\Delta\nu/\nu < 10^{-4}$ and for $B > 1680 D$ we need (3.75) instead of (3.75a). In practice integration times smaller than τ are often used, which means that actual numbers of samples differ from the (3.75) or (3.75a), which means that for results using these equations we need to state the proper conditions and regimes for which they are indeed valid.

3.7.1.2 2-D FFT facet imaging

The processing volume C_{FFT} for an FFT with N_p pixels equals

$$C_{FFT} = \frac{1}{2} N_p \log_2(N_p) \quad [CMA]$$

A set of N_f small 2-D facet FFTs requires $\frac{1}{2} N_f (N_p/N_f) \log_2(N_p/N_f) = \frac{1}{2} N_p (\log_2(N_p) - \log_2(N_f))$ CMA. Hence, a single large 2-D FFT requires about the same processing capacity as a set of small ones that provides the same total number of image pixels since $\log_2(N_f) \ll \log_2(N_p)$.

The grid spacing assumed in section 3.4.2 leads to a field diameter $3 \lambda/D$ and with sampling of 3 pixels per resolution element of width λ/B_{max} we get for the extent N_e of the FFT grid in pixels

$$N_e = 9 B_{max} / D$$

The total number of pixels N_p follows from $N_p = N_e^2$ and leads to a total FFT processing capacity of

$$C_{FFT} = 40 (B_{max} / D)^2 \log_2(40 (B_{max} / D)^2) \quad [CMA]$$

For estimation of $\log_2(40 (B_{max} / D)^2)$ we take $(B_{max} / D) \sim 10^4$ which is representative but not very critical to find ~ 32 leading to

$$C_{FFT} = 1280 (B_{max} / D)^2 \quad [CMA] \quad (3.76)$$

The additional attenuation over the field introduced by the convolution operation needs correction after transformation as discussed in section 3.3 and only the centre quarter of the field is retained for further processing.

3.7.1.3 Number of Facets and Size of the Convolution Kernel

In subsection 3.1.8 we derived (3.25) to find the radius $\Delta\theta_r$ of the FoV of a 2-D FT image for a non-planar array with a maximum phase error π^{-1} caused by 2nd order terms. A station with aperture diameter D and parabolic taper has a beam with HWHM $\sim 0.64 \lambda / D$ and conventional polyhedron imaging requires a number of facets N_f^P given by

$$N_f^P = (0.64 \lambda / D)^2 / \Delta\theta_r^2 \quad (3.77)$$

Evaluation of $\Delta\theta_r$ according to (3.25) using $W = B_{max} / 2\lambda$ gives for the polyhedron case

$$N_f^P = 1.8 \lambda B_{\max} D^{-2} \quad (3.77a)$$

In subsection 3.4.5 we derived the size of the FoV as function of the non-planarity for a maximum phase error of π^{-1} rad by higher order terms in case 2nd order terms are corrected by a complex convolution. We found different equations depending on the choice for the reference plane for the U,V-coordinates from which follows (3.50b) for the minimum aperture diameter D_{\min}^E of a facet beam in the extrinsic case and (3.50c) for D_{\min}^I valid for the intrinsic case. These minimum aperture diameters lead to a maximum extent of the convolution kernel that has to correct for 2nd order terms, which in turn could drive processing requirements beyond what is affordable.

Instead of correcting for a full station beam that limits the maximum baseline, or even a maximum facet beam that minimizes the number of facets for a given maximum baseline we now ask for the number of facets needed when we use the smallest complex convolution kernel. This situation minimizes the total processing since Fourier processing is almost invariant for the number of facets.

The extent of the convolution kernel is given by (3.45) but we use aperture diameter $D_f > D$ that defines the number of facets beams within a station beam. We have two situations, one for extrinsic and one for intrinsic coordinate configuration. The number of facets N_f that fill the centre of the station beam within half power is for the two cases given by

$$N_f^E = (D_f^E / D)^2 \quad (3.78)$$

And

$$N_f^I = (D_f^I / D)^2 \quad (3.78a)$$

Deriving D_f^E from (3.78) and inserting it in (3.45) using $H = B_{\max} / 2$ gives

$$\begin{aligned} K_f^E &= (7 / N_f^E) (\lambda / D) (B_{\max} / D) \\ &= (7 / N_f^E) (\lambda / B_{\max}) (B_{\max} / D)^2 \end{aligned} \quad (3.79)$$

Deriving D_f^I from (3.78a) and inserting it in (3.45) using (3.26) to find H by Earth curvature over distance L_{\max} and assuming $L_{\max} = B_{\max} / 2$ for a symmetric array configuration with diameter B_{\max} we find

$$K_f^I = (7 / N_f^I) (\lambda / 4R_E) (B_{\max} / D)^2 \text{ for } B_{\max} < R_E \quad (3.79a)$$

with Earth radius $R_E \sim 6,371$ km. Both formulas show proportionality to the FoV of the station beam expressed in area resolution elements. However, there is a dramatic reduction factor $B_{\max} / 4R_E$ since we need to correct only for intrinsic non-planarity instead for extrinsic non-planarity as given by the ratio of maximum baseline over four Earth radii.

In practice there is a minimum diameter of 7 pixels for the linear extent of the convolution kernel to get sufficient accuracy, which leads to different minimum numbers of required facets N_f^E for extrinsic and N_f^I for intrinsic convolution correction, respectively, given by

$$N_{fmin}^E = \lambda B_{max} D^{-2} \quad (3.80)$$

and

$$N_{fmin}^I = \lambda B_{max} D^{-2} (B_{max} / 4R_E) \quad (3.80a)$$

Interestingly, the number of facets in (3.80) is equal to the number of planes in 3-D FT imaging as defined by (3.11) but the total FFT processing of all facets together is even smaller than for a single plane in 3-D FT imaging as explained in 3.7.1.2. In contrast dealing with only intrinsic non-planarity requires according to (3.80a) much less facets but the duration of a synthesis observation is limited due to the choice of coordinate system and could require additional sets of 2-D facet images. Each set has however about the same size as a single plane in the 3-D approach.

A somewhat disappointing result is that N_{fmin}^E is only a factor 1.8 smaller than N_f^P if we use the minimum kernel size of 7^2 pixels. A major difference is that in the polyhedron case we have a maximum phase error of π^{-1} at the half power of each facet beam by 2^{nd} order terms while the convolution correction leaves only 4^{th} order terms that are much smaller. However, we can reduce N_{fmin}^E by making $K_f^E > 7$ at a progressively increasing processing penalty.

A large number of facets is needed for long wavelength and small stations. For Dutch LOFAR we have the option to vary the size of an LBA station by selecting any subset of 48 antenna elements from 94 that can actually be combined in a station beam. In this way we are able to realize at each frequency a maximum FoV at full sensitivity. The worst case situation is reached with the 32 m configuration at 50 MHz and we give the minimum number of facets as function of baseline in table 3.3 for extrinsic non-planarity and intrinsic non-planarity correction. We give two lines for extrinsic baselines, one using the conventional real kernel resulting in the number of facets needed for conventional Polyhedron imaging and one for a complex kernel.

Polyhedron imaging has a phase deviation that scales with the square of the distance from the field centre and reaches for the longest baseline a maximum phase deviation of π^{-1} rad for a source at half power. With a complex Gaussian convolution these deviations are corrected but there remain small terms of fourth order that are not corrected. For the intrinsic situation 3^{rd} order terms are left proportional to the duration of a synthesized snapshot given by (3.56) as discussed in subsection 3.5.3. Eliminating the facet size $\Delta\theta_{cr}$ from (3.56) by inserting $\Delta\theta_r$ according to (3.77) makes $|\Delta I|$ dependent on N_f^P .

Inserting (3.80a) shows that the tracking range stays limited to ~ 10 min, independent of the number of facets.

The kernel size for convolution has been taken to minimize the required processing. More facets could be made for instance to facilitate corrections for direction dependent effects without introducing an additional processing penalty for imaging. Less facets could be needed if the convolution kernel is extended, which requires progressively increasing processing.

In table 3.4 we give results for the HBA stations of LOFAR that work at higher frequencies than the LBA stations.

The number of facets could be decreased but the linear extent of the convolution kernel increases inversely proportional driving up convolution processing with the square. This option might be attractive for line imaging where fewer visibilities are processed per image and where Fourier inversion remains the dominating processing per image.

Table 3.3. Minimum number of facets for 32 m LBA stations at 50 MHz using a 7^2 convolution* kernel

Baseline	3 km	30 km	60 km	90 km	300 km	600 km	1200 km
Extrinsic R*	34	317	634	951	793**	1585**	3170**
Extrinsic C*	19	176	352	528	440**	880**	1760**
Intrinsic C*	1	1	1	4	22***	88***	88**

* Real and Complex convolution kernels.

Intrinsic images are limited to ~ 10 min duration and get max π^{-1} phase error at half power by tracking.

Extrinsic images have no tracking limitation and much lower residual 4th order phase errors.

** Baselines of 300 km and longer contain European stations that have a station diameter of 65 m and a reduction by a factor 4 has been included to cover just the centre half of the smaller station beam.

*** The 300 and 600 km baselines are formed by stations at 300 and 600 km from the centre of the array and also the factor 4 reduction has been included.

3.7.1.4 Fast Faceting

We have seen that the fast faceting algorithm can provide a number of facet datasets, where smaller facets have fewer samples with longer integration time and larger bandwidth, but the total amount of samples in all facets together is constant.

The most important aspect of fast faceting is that by increasing the number of facets the linear extent K of the convolution kernel can be reduced. Although the total amount of samples that need to be convolved is constant, the total processing volume for convolution is reduced with a smaller kernel. If faceting is continued beyond the level where the convolution kernel reached its minimum practical linear extent of 7 pixels, there is no further processing advantage. However having more facets could still be attractive for facet based calibration and correction approaches and processing capacity could be saved if only a part of all available facets is kept for imaging and analysis. A disadvantage is that the number of U,V-samples in each facet image is reduced, which could potentially increase the average side lobe level of the psf.

Table 3.4. Minimum number of facets for 40 m HBA stations at 150 MHz using a 7^2 convolution* kernel

Baseline	3 km	30 km	60 km	90 km	300 km	600 km	1200 km
Extrinsic R*	7	68	135	204	169**	338**	675**
Extrinsic C*	4	38	75	114	94**	188**	376**
Intrinsic C*	1	1	1	1	4***	18***	18**

* Real and Complex convolution kernels.

Intrinsic images are limited to ~ 10 min duration and get $\max \pi^{-1}$ phase error at half power by tracking.

Extrinsic images have no tracking limitation and much lower residual 4th order phase errors.

** Baselines of 300 km and longer contain European stations that have a station diameter of 65 m and a reduction by a factor 4 has been included to cover just the centre half of the smaller station beam.

*** The 300 and 600 km baselines are formed by stations at 300 and 600 km from the centre of the array and also the factor 4 reduction has been included.

3.7.1.5 Minimum number of convolution operations

The minimum number of convolution operations is reached when the facet size is decreased to the level where the required convolution kernel reaches its minimum linear extent of 7 pixels. Convolution of a single complex visibility datum to 7^2 pixels on the square grid for the 2-D FFT facet image needs 49 CMA operations, so the minimum processing volume C_{cm} needed for convolution for a continuum image with $\Delta v > \delta v$ is given by

$$C_{cm} = 49 N_{sa}^c = 0.11 N_b (\Delta v / v) (B_{max} / D)^2 T_s \quad [\text{CMA}] \quad (3.81)$$

We see that the minimum processing capacity for convolution in 2-D Facet Fourier imaging is proportional to $(B_{max} / D)^2$ i.e. to the total FoV expressed in area resolution elements, to relative bandwidth, to total time and to total number of baselines.

The convolution kernel depends on the actual W-value of each observed U,V-sample. This means that for every U,V-sample along a track we need to introduce a small modification in each kernel element. Assuming some linear interpolation we just take 2 CMA per kernel element instead of 1 CMA and find

$$C'_{cm} = 0.22 N_b (\Delta v / v) (B_{max} / D)^2 T_s \quad [\text{CMA}] \quad (3.81a)$$

If the facet size is however increased in an attempt to reduce the total number of facets below the minimum number defined by (3.77) we pay a processing penalty since the cost of convolution increases. In case of short observations with a few baselines the penalty might be acceptable compared with the processing required for the FFT. In such a case a direct Fourier transform might be the most effective solution.

For 3-D FFT imaging in principle a 3-D convolution kernel is needed. In practice the 3-D transformation is done per 2-D plane and we require 49 CMA per baseline sample for each plane. Actually we still have a 3D convolution, but with a top hat with a width equal to the distance between planes in the 3rd dimension.

3.7.1.6 Number of source subtract operations

Hybrid imaging requires that the strongest sources are subtracted from the visibilities such that imaging artefacts of all remaining sources create only minor additional side lobes to point sources in a Fourier image. This requires an accurate system model with parameters that describe the complex station gain at every position in the station beam. In chapter 4 we will discuss the accuracy of these parameters and the resulting errors in the final image will be discussed in chapter 5.

For each point source to be subtracted we calculate the complex gain factor for each baseline as the product of two station factors. Subtraction from the complex visibility requires an additional CMA so 3 CMA per source subtract. It has been verified that the subtraction procedures developed for the LOFAR calibration package indeed perform according to this estimate and that overhead in calculation of station positions can be neglected compared to the large number of spectral channels per baseline per station.

For the visibilities of point sources inside a facet and adjacent facets there is only a small additional decorrelation factor that depends on the square of the phase change per baseline over integration time and over bandwidth as discussed in section 3.2. This additional correction can therefore also be derived from station based phase changes over integration time and spectral bandwidth. For a phase change less than $\pi/2$ per station we need in principle 3 real multiply add operations, which we count for 1 CMA. For objects outside this limited area a full amplitude correction is needed based on more accurate evaluation of the sinc function for which we assume the equivalent of 3 CMA. In practice this applies to a very limited set of ~ 10 objects outside the main beam of the station, which require an accurate source model.

From the previous reasoning it is clear that subtraction of only 10 sources requires already the same processing volume as the convolution operation. In chapter 4 we will see that we need at least 5 strong calibration sources inside the station beam to do reasonable imaging at all and these have to be subtracted using a simple sinc evaluation. We conclude therefore that we need for subtraction at least the same number of CMA operations as required by the kernel size discussed in the previous heading.

3.7.1.7 Station beam and polarization correction

Polarization corrections are based on simple average gain corrections given by the coefficients of the Mueller matrix per observing interval of order 10 min as discussed in section 3.6 and could equally well be applied to observed polarization coherences as to coherence images provided by a short synthesis image. A simple gain slope over an image would however require a more complicated convolution kernel but the number of operations does not change as long as a linear extent of 7 pixels suffices. In the latter case even station dependent corrections could be applied just as for the non-planarity.

We need only one real multiplication per pixel for each of the four observed coherence images to produce an image in one Stokes parameter. This is the equivalent of 1 CMA per pixel in a final image and needs to be performed over only a part of the full FFT images and can therefore be neglected when compared with the 32

CMA per pixel in the full FFT. Of course all 4 coherency images need to be made even in case only one Stokes parameter would be finally needed.

3.7.1.8 Interpolation on a sky image grid

In case we form a synthesis observation from a number of synthesized snapshot images every snapshot pixel needs to be integrated on the appropriate sky grid point. Such an interpolation requires a gridding kernel with a linear size that is at least equal to the number of grid points per resolution beam. Assuming a 3x3 interpolation, we need only 9 Real Multiply Add operations per pixel and we use only the central $\frac{1}{4}$ of the snapshot image then. Also this operation can be neglected when compared with the FFT operation for each segment.

For the 3-D imaging we also need interpolation of the image planes on a final sky sphere, which is an operation comparable to the segment interpolation and can also be neglected when compared with the FFT operation.

3.7.1.9 Number of synthesized 2-D FFT snapshots and number of planes in 3-D

In section 3.5 we derived a minimum duration for a synthesized 2-D FFT snapshot image that depends on the size of a facet field fixed on the rotating sky when observed in an Earth bound array coordinate system at fixed azimuth and zenith angle. Although the field rotation during the segment synthesis is corrected by rotation of the U,V-coordinates that are first corrected for projection, there is no need for polarization rotation correction if the tracking time is limited to 10 min. In section 3.6 we have shown that tracking of the sky field with a phased array station beam through the beam of the average element antenna causes instrumental polarization effects. It was also shown that so called bi-scalar self-calibration on an un-polarized source in the sky field leads to first order polarization correction for the whole sky field viewed by the station beam. Additional corrections for changes over each facet in the station beam could be applied either to the visibility data or to the segment image. Corrections for differential Faraday rotation between two station need however to be made using the four polarized visibilities of each baseline. An average Faraday rotation correction for all stations together could be sufficient if the tracking time is limited to 10 min. Also effects of fine structure in the average element beam such as blind angles are reasonably averaged out over such a tracking time interval.

The synthesized snapshot approach makes for every interval of about 10 min a set of facet images each in a dedicated coordinate system that allows simple correction for the non-planarity of the array. The individual facet images need to be interpolated and integrated on a common sky based coordinate grid which allows straight

forward correction for field distortions such as differential refraction effects induced by the ionosphere. At the same time the effects of varying shape and polarization properties of the beam of a phased array antenna station that tracks a field in the sky could be corrected as well.

The number of synthesized snapshots N_{ss} is determined by the total synthesis time T_s and the synthesized snapshot duration $T_{ss} \sim 600$ s and follows from

$$N_{ss} = T_s / T_{ss}$$

For 3-D imaging we need the same type of corrections to the image field as for the synthesized snapshot approach. However, since all baseline data of a long synthesis are transformed with a single 3-D FFT we can apply corrections only by a multiplication operation per visibility or by a convolution operation per visibility. We assume that the 3-D convolution for these corrections can still be described by a 7^2 convolution per baseline sample for each plane.

Equation (3.10) gives the number of planes N_{pl} in the 3-D FFT:

$$N_{pl} = \lambda B_{max} D^{-2}$$

3.7.2 Balancing Convolution and source subtraction against FFT processing

We have shown that the processing volume required for convolution, source subtraction and FFT imaging are all three proportional to $(B_{max} / D)^2$ i.e. to the number of resolution elements in the total FoV. Also we have seen that there is a minimum number of sources that need to be subtracted accurately which requires a processing volume that is about equal to that required for convolution with a limited kernel size.

The minimum kernel size minimizes not only the processing time for convolution but also defines the maximum facet size that can be imaged with a 2-D FFT for a given maximum non-planarity and a maximum tolerated phase deviation of π^{-1} . The number of facets that are needed to cover the full station main beam are given by (3.80) or (3.80a) for extrinsic or intrinsic non-planarity correction respectively. However, the number of facets has only minor impact on the total capacity required for FFT processing of a given total FoV.

The minimum ratio $R_{minV/I}$ of visibility related processing over image related processing is related to (3.81a) and (3.76). We find for subtraction processing equal to convolution processing

$$R_{minV/I} = 2 C'_{cm} / C_{FFT} = 3.5 \cdot 10^{-4} N_b T_s \Delta v/v \quad (3.82)$$

For a typical long synthesis observation we have $T_s \sim 20,000$ s and we get

$$R_{\min V/I}^{\text{Long}} = 7 N_b \Delta v/v \quad (3.83)$$

These two equations compare the processing of a single large FFT (or a set of smaller facet FFT's) with the minimum processing for the total number of visibilities used in that image for a given relative bandwidth of the visibilities.

3.7.2.1 Source Subtraction dominates over convolution and Fourier inversion

Processing requirements for subtraction of at least the 10 strongest sources in the station main beam and side lobes are about equal to processing requirements for convolution with a limited complex kernel size. Current practice for LOFAR requires subtraction of an additional $\sim 10^3$ sources in the main beam to reach the thermal noise in a 6 h continuum image, which demonstrates that the processing for imaging is completely dominated by source subtraction. Chapter 5 will analyse whether such a large number of source subtracts is still necessary if accurate hybrid imaging is used in combination with multi direction self-calibration that provides an appropriate phase screen for accurate subtraction. For the moment we assume that in practice we need subtraction of at least an additional set of 100 sources in the main beam and the ratios given by (3.82) and (3.83) increase by a factor ~ 5 leading to:

$$R'_{\min V/I} = 2 C'_{\text{cm}} / C_{\text{FFT}} = 17.5 \cdot 10^{-4} N_b T_s \Delta v/v \quad (3.82a)$$

$$R'_{\min V/I}^{\text{Long}} = 35 N_b \Delta v/v \quad (3.83a)$$

3.7.2.2 Continuum versus line observing

Most line imaging applications have after subtraction of the dominant continuum contribution a resulting image with low signal to noise ratio, which means that disturbing effects of error side lobes by insufficient calibration do not lead to observable degradation of the noise in an image. This means that line imaging needs no additional source subtracts. Instead of (3.82) that used (3.81a) we need (3.75a) to derive the equivalents of (3.82) and (3.83).

For continuum observing with $\Delta v/v > 3 \cdot 10^{-3}$ we are according to (3.83a) visibility processing dominated in a 6 h observation using only 10 baselines. According to (3.80) we need $B_{\max} < D^2/\lambda$ to cover the full station beam with a single FFT using a 7^2 complex convolution kernel.

The number of baselines in an array with N_{st} station is given by

$$N_b = \frac{1}{2} N_{st} (N_{st} - 1)$$

Insertion into (3.82a) shows that an observation of about 10 minutes would need ~27 stations to become visibility processing dominated. Although a long synthesis observation needs a synthesized snapshot image every 10 min, it stays visibility processing limited even for the smallest complex convolution kernel assuming that at least 100 sources have to be subtracted in a continuum image with $\Delta v/v > 3 \cdot 10^{-3}$ with an array using more than 27 stations, such as LOFAR.

3.7.2.3 3-D and 2-D synthesized snapshot imaging alternatives

Facet imaging reduces a single large FFT into a set of smaller ones that require together even less processing and minimizes the processing capacity for imaging. It has however to be realized that working with facets needs a dedicated data organization which has impact on the design of an imaging package that has to run on a HPC platform. Moreover, the facets have to be stitched together, which requires additional processing. Instead, when there are more than 27 stations visibility processing already dominates a 10 min observation and there is no serious processing penalty if the FFT processing is increased either by more planes in the 3-D imaging approach or by more synthesized snapshot images in a 2-D approach. The main reason to accept such relative minor increase in total processing time is that stitching of facets could be avoided. However, the alternative 3-D imaging requires unfortunately an interpolation to get a single image with associated artefacts.

In view of the ~ 400 Fourier planes for 3-D imaging with LOFAR, this option can be ruled out in favour of facet imaging with baselines up to 1200 km. However, for observations with limited duration using stations closer to the core than 80 km, as for the Dutch sub array, faceting can be avoided for continuum imaging with 0.3 % relative bandwidth per image using the synthesized snapshot approach.

3.7.3 Comparing post correlation processing with correlation processing

An important aspect in the design of synthesis arrays relates to the distribution of the total collecting area over a number of antenna stations such that appropriate resolution is obtained in a large FoV but foremost that the thermal noise as determined by collecting area, bandwidth and integration time is indeed realized over that FoV. The main problem is reducing imaging artefacts to the level where they are indeed smaller than the thermal noise. This should be done against acceptable cost of the associated processing.

Acceptable cost relates to other costs that are also unavoidable such as the cost for antenna stations and the cost for signal reception, transport and cross-correlation. Optimization of the cost distribution is possible depending on observing strategy

[Bregman, 2004a] and leads to the cost of cross-correlation and imaging together of less than about 15 % of the total instrument cost. It is therefore relevant to compare the total cost of imaging with the total cost of correlation.

In terms of system design we need the imaging to keep up with correlation to avoid loss of data. When correlation platform and imaging platform are matched in average output data rate and average input data rate respectively, this simplifies the comparison of the platform requirements even further. Instead of volume comparison in CMA by the processes that run on the platforms we can compare processing power of the platforms expressed in CMA/s which relates to flop/s where a flop is a floating point operation such as a multiply or an addition and we assume 6 flop for 1 CMA. The minimum processing power P_{C+S} for convolution and subtract together follows from (3.81a)

$$P_{C+S} = 12 C'_{cm} / T_s = 2.6 N_b (\Delta v / v) (B_{max} / D)^2 \quad [\text{flop/s}] \quad (3.84)$$

It has been demonstrated that source subtraction on a PC type platform is processing limited and not by internal data transport. A cluster of PC type platforms can indeed handle the many parallel data streams from the correlation platform [Schaaf, 2004], [Schaaf, 2003] and the total performance can therefore be expressed by the processing power of the sum of all elements in the cluster.

The processing power of a so called FX cross-correlation system as used by LOFAR is dominated by the CMA operations needed for cross-correlation where samples from the data streams of each telescope pair are pair wise multiplied and integrated to a complex visibility sample [Romein, 2010]. Various processing platforms have been compared [Nieuwpoort, 2009] but only IBM's BlueGene/P allows full use of its peak processing power for correlation. A correlation system that provides N_b visibility samples with total bandwidth Δv requires a processing power P_{cc} given by

$$P_{cc} = 6 N_b \Delta v \quad [\text{flop/s}] \quad (3.85)$$

The ratio $R_{C+S/cc}$ of minimum convolution plus minimum source subtract processing power over cross-correlation processing power is now given by

$$R_{C+S/cc} = (0.43 / v) (B_{max} / D)^2 \quad (\text{with } v \text{ in Hz}) \quad (3.86)$$

This ratio expresses the processing power ratio for the two platforms that each have a HPC architecture optimized for the specific purpose and have sufficient processing power, memory, internal data routing capacity and data rate for external input and output. In case the price of a dedicated platform is dominated by processing elements of comparable technology, the ratio $R_{C+S/cc}$ defines the cost ratio of the post correlation platform over correlation platform. This assumes that adver-

tised processing power of the different platforms is indeed the realized value when the respective applications are performed.

Interestingly, this ratio of processing powers is independent of the actual bandwidth that will be used in a single synthesis image and leads to simple conclusions when imaging is indeed dominated by visibility processing as is the case for LOFAR and larger arrays as the SKA. Although the total number of baselines and the total bandwidth per baseline only determine the processing power of a platform performing FX correlation, the required output data rate is proportional to the number of resolution elements in the effective FoV. This effective FoV is only a fraction of the total FoV offered by the station beam and can be portioned in smaller data sets for facets centred on a region of interest. The total number of facet datasets defines the input data rate of the imaging platform and its processing power for continuum imaging that is proportional to this data rate and to the number of sources that need to be subtracted. In fact, the processing power for convolution is the same as for subtraction of 10 sources and together they dominate over Fourier processing. In practice even more sources need to be subtracted, making it the dominant post correlation processing activity for continuum imaging.

3.7.3.1 Continuum imaging dominated by correlation

LOFAR with 32 metre stations operating at 50 MHz would require for 120 km baselines $R_{C+S/cc} \sim 0.12$ but 1200 km baselines would require $R_{C+S/cc} \sim 12$ if the full FoV of a small station needs to be imaged in real time. Fortunately the European stations have 68 m diameter requiring $R_{C+S/cc} \sim 2.7$. The large ratio relates to the 1.7 % decorrelation by bandwidth and by time integration that is tolerated on the longest baseline for a point source at half power of the facet beam, which drives up N_{sa} defined by (3.75) and used in (3.81).

As discussed in section 3.4.5 this tolerance could be relaxed to 7 % which reduces the amount of data given by N_{sa} in a facet dataset by a factor 4 and the processing power for convolution accordingly. Even with this reduction not all facets over the station beam can be imaged by an imaging platform that has typically a factor 10 lower processing power than the correlation platform.

The most important conclusion at this stage is that making continuum images with Dutch LOFAR is dominated by the processing power required for cross-correlation, and indeed the biggest HPC platform is dedicated to this task. Issues related to long baselines such as fast faceting and data rate between correlation platform and imaging platform have already been discussed in section 3.4. However, if substantially more than 10 sources have to be subtracted to reach the thermal noise floor in a continuum image the processing power of the imaging platform has to be increased accordingly. The required number of these additional sources that have to be subtracted will be discussed in chapter 5.

3.7.3.2 *Comparison with conventional imaging packages and their successors*

Processing estimates for legacy image forming packages that are said to be dominated by convolution processing range up to 30,000 flop per spectral visibility sample. Detailed simulations using an early version of the W-projection method [Cornwell, 2004] show that 10,000 flop could be considered as programming overhead leaving 20,000 flop for convolution and source subtraction. We arrived at ~100 CMA or ~600 flop for convolution dominated imaging and ~360 flop for subtraction of the 10 strongest sources in a field. Apparently in these packages either too many processing cycles are used in the convolution or ~500 more sources are subtracted from the measured data that indeed contained 250 sources. In view of the iterative processing approach that repeats every step we conclude that our simple analysis is confirmed by independent practical implementation. The need for subtraction of a large number of sources will be investigated in chapter 5.

In a recent analysis [Humphreys, 2011] a convolution kernel with linear extent of 65 pixels is proposed for spectral line imaging to handle extrinsic non-planarity of baselines up to 2 km. With W-projection the full beam of a 12-m station observing at 0.2 m wavelength could be handled. For imaging with 6 km baselines a kernel extent of 110 pixels would be needed. These numbers produce good imaging results and are indeed consistent with our derivations. However, for 36 stations, line imaging is no longer Fourier processing dominated after ~3 min, which suggests that the synthesized snapshot imaging approach requires less processing. However, conventional imaging using only 8 facets and the smallest complex convolution kernel is Fourier processing limited for a 6 h observation.

Correlation [Romein, 2010], [Bregman, 2010] and convolution [Humphreys, 2011] could indeed be realized on different appropriate types of processing platforms that according to their specifications should provide the appropriate mix of resources to maximize throughput for the specific application.

3.7.4 **Results and Conclusions**

Our comparison of the processing required by different imaging approaches yields the following results:

- Single facet 2-D FFT imaging with complex quasi-convolution correction for extrinsic non-planarity is limited to line applications with limited baseline extent, where FFT processing dominates.
- Imaging with longer baselines requires a larger linear kernel extent and makes convolution the dominant processing that increases quadratic with maximum baseline.

- Faceted 2-D FFT imaging subdivides a large FFT into a number of smaller transforms, but total processing is practically proportional to total number of pixels. Therefore, Fourier transformation is proportional to the total number of resolution elements spanned by the beam of the stations in a synthesis array, irrespective of number of facets.
- The number of visibility samples in a continuum image is also proportional to this number by the requirement that bandwidth and integration time smearing need to be avoided. Quasi-convolution processing dominates in all practical situations over Fourier transformation when a minimum-size, complex kernel is used.
- Continuum imaging with more than 0.3 % relative bandwidth by an array with more than 27 stations is dominated by source subtraction over Fourier transforming for integration longer than 10 min, irrespective of FoV and resolution if more than 100 sources need to be subtracted.
 - This makes a long synthesis as a sum of a number of synthesized snapshot images a feasible approach from processing perspective.
- Synthesized snapshot imaging uses a single large 2-D FFT but tracking of a source far from Zenith practically limits the duration of a synthesized snapshot image to order 10 minutes.
- A long synthesis image therefore needs a number of shorter observation for which the images need appropriate coordinate conversion before integration to a full image in sky coordinates.
- 3-D FFT imaging makes a set of 2-D quasi-images that need to be interpolated on a single spherical surface, but uses a small and real convolution kernel.
- 3-D FFT imaging could be attractive for continuum imaging where convolution and source subtraction could still dominate the post correlation processing at sufficiently high frequencies.
- Processing estimates for conventional image forming packages that are said to be dominated by visibility processing range up to 3,000 CMA per spectral visibility sample. We predict only 100 CMA for convolution and 6 CMA per source subtract, which suggests that many more than a minimum of ~20 sources are usually subtracted. This is investigated in chapter 5.

3.8 Summary, Results, and Recommendations

In this chapter on wide field synthesis imaging a study is presented that identified the limitations of conventional imaging approaches for application by LOFAR and SKA. A comparison is made between 3-D Fourier inversion, approximate 2-D Fourier inversion for a large set of small facet images also called polyhedron imaging, and approximate 2-D Fourier imaging after complex convolution correction of the non-planar baselines also called W-projection. All three methods use convolution to a rectangular grid of the observed correlation data of all pairs of telescopes that form together the synthesis array followed by Fast Fourier Transformation. This combination of convolution and FFT forms the basis for efficient imaging.

The introduction of this chapter outlined some of the results that motivated the work. In six sections, we addressed the various aspects and we concluded each section with a summary of specific results, of which most are known but these serve as the context for the new results

In section seven, we combined the results of previous sections. We found that the processing for image forming is minimized by subdividing a large FoV in a number of smaller facets that need a small kernel for their convolution to a rectangular grid. The typical minimum size for such a kernel is 7^2 and a complex kernel could correct 2nd order terms in the expansion of the non-planarity factor and maximizes the FoV of the facets. For such minimized processing for image forming, we found three fundamental scaling relations:

- Fourier transformation by FFT processing is almost independent of the number of facets but proportional to the total number of resolution elements in a required FoV, i.e. proportional to $(B_{\max} / D)^2$ where B_{\max} is the maximum distance between the stations and D the diameter of the stations in a synthesis array.
- The number of visibility samples for a facet beam of a continuum image is given by the relation $f(N_b T_s \Delta v/v) (B_{\max} / D_f)^2$, where N_b is the total number of baselines, T_s the duration and $\Delta v/v$ the relative bandwidth of the synthesis observation, while D_f is the aperture diameter for the facet beams.
 - This relation assumes that the total bandwidth Δv is covered by a number of channels such that delay smearing is avoided just as integration time smearing and this requirement introduces the factor $(B_{\max} / D_f)^2$.
 - The factor $(N_b T_s \Delta v)$ determines the sensitivity of an image and suggests working with fewer but larger stations. In chapter 5 we will analyze the relation between N_b and the number of sources that need to be subtracted accurately to reach the thermal sensitivity. If this number of sources exceeds 10 the processing re-

- requirements for subtraction will dominate the complex convolution processing.
 - The proportionality constant f is determined by the tolerated loss in sensitivity on the longest baseline for sources at half power of the facet beam. This loss is proportional to the square of the radius of the FoV, and in practice a maximum of 1.7% is chosen but visibility processing could be reduced if larger losses are accepted.
- Observing with 0.3 % relative bandwidth using an array with more than 27 stations spanning baselines longer than 1.5 km provides a large number of visibility samples per unit time. Subtraction of 100 sources and quasi-convolution processing with the smallest kernel will dominate over Fourier processing after 10 min, making synthesized snapshot imaging a viable option for continuum imaging with LOFAR and SKA.

These results are dramatically different from the often-used formula valid for 3-D Fourier imaging that includes an additional processing penalty factor $\lambda B_{\max} / D^2$ also called Fresnel factor. This factor accounts for the number of 2-D Fourier planes that are needed to provide adequate FoV for a 3-D baseline distribution, which results if a 2-D array tracks a sky source for some time. The formula indicates that the 3-D imaging approach might not be attractive for wide field imaging at low frequencies but 2-D alternatives have their own problems. Polyhedron imaging for instance suffers from a large number of facets that is at least equal to the number of planes in the 3-D case. This would itself not drive up the total cost for Fourier processing, but current implementations lack the fast faceting method that reduces the number of visibilities per facet. More serious is that still a relatively large phase error has to be tolerated on the longest baselines for sources near the edge of a facet. As shown, complex quasi-convolution with a small kernel could reduce this error substantially.

Such a convolution using a larger kernel, as in the proposed W-projection method, could potentially correct a FoV as large as a station beam. We have given proof that Gaussian quasi-convolution corrects for second order terms in the expansion used for the non-planarity factor in a 2-D Fourier transform for a non-planar array. However, this proof has not been extended to higher order terms that could possibly be corrected by a non-Gaussian quasi-convolution.

Unfortunately, second order correction of the full beam of a 32 m LOFAR station at 50 MHz would allow a maximum baseline of only ~7 km. In addition, the associated Gaussian quasi-convolution processing would be greater than the processing required for correlation due to the large size of the required complex convolution kernel, which is not an attractive option from the perspective of system optimization.

Apparently, a combination of faceting using the proposed fast faceting approach together with a small but complex kernel for quasi-convolution is a solution that minimizes the total processing load and provides low image distortion by not corrected higher order terms. However, a single facet approach could be attractive for line observing when fewer visibilities allow a larger kernel, but this is only feasible for compact arrays with small non-planarity, and more importantly, with narrower beams as is the case at much higher frequencies than 50 MHz, such as for instance for ASKAP.

We analysed Fourier imaging from first principles and found that current imaging packages concentrate on solving the issues related to extrinsic non-planarity. These issues are in fact a problem created by the particular choice of the coordinate system with its W -axis towards the centre of the field of interest that allows simplifying approximations for non-coplanar baseline effects. Even an intrinsic planar array suffers from extrinsic non-planarity effects when a sky source is tracked, since its plane tilts with respect to the plane used for 2-D Fourier imaging. However, a planar array allows accurate hemispheric imaging of a stationary sky by 2-D Fourier inversion, in a coordinate system with its W -axis towards local Zenith.

Intrinsic non-planarity in a quasi-planar array with stations that follow Earth curvature is much smaller than extrinsic non-planarity caused by projection of baselines on the direction towards the source field. An array based coordinate system would therefore in principle need either a much smaller convolution kernel or a much smaller number of facets, if source movement could be ignored as in snapshot imaging.

A full analysis has been presented that handles the consequences of 2-D Fourier imaging using fringe tracking of a shifting and rotating sky field in a coordinate system that uses a reference plane with minimum distances to all stations. This analysis has shown that wide field 2-D Fourier inversion is well possible for a quasi-planar array where stations have small deviations from a reference plane. However, a sky field with substantial FoV at nonzero zenith angle, requires telescope based corrections for first order phase deviations of the observed visibilities such as coordinate projection and rotation and second order corrections by quasi-convolution. When these corrections are applied, third order effects limit the duration of a synthesized snapshot observation depending on size and elevation of the FoV. This third order phase term is not only proportional to non-planarity and to the cross product of distances between source and field centre in both image coordinates but becomes third order since there is also proportionality to the required shift between actual sky field and image centre defined for the middle of the tracking interval. The true limitation of the synthesized snapshot method is this third order term that determines the maximum allowed tracking time for a given FoV size. This size is practically defined by the maximum tolerated phase error on the baseline with the largest non-planarity. However, such a 3rd order term does not exist for 2-D Fourier

imaging with W-axis towards the centre of the field, since that method has only 2nd and 4th order terms.

Synthesized snapshot images of ~10 min need subdivision of the FoV in a number of smaller facets when remote LBA stations of 32 m diameter are included that are further than 80 km from the core of the array. In that case, the facet extent is effectively determined by the 3rd order phase term when a tracking range is specified smaller than the radius of the facet. Second order terms need to be corrected by a complex convolution kernel of adequate size. The 3rd order phase deviations in the baselines with the largest non-planarity needs to be less than ~0.3 rad for an object at half power level of the facet beam. This phase tolerance leads to about 1.7 % degradation in the visibility on some baselines and is comparable to the degradation introduced by the finite integration time and bandwidth in the visibility on the longest baseline for objects at half power level of the facet beam. This degradation leads to an additional effective taper of the visibilities for these sources which results in broadening of these sources, which complicates deconvolution when a nominal psf is used for the whole facet field.

Results

In subsequent paragraphs we summarize the new results of this work.

Our first analysis derived a relation for the minimum diameter of the complex kernel used for quasi-convolution of observed visibility data that properly corrects for second order effects in 2-D Fourier imaging with non-planarity of specific baselines. This analysis showed that a kernel size of 7^2 would be adequate for stations up to 80 km from the core if only non-planarity due to Earth curvature needs to be corrected to provide correction for the full FoV of a station of 32 m diameter observing at 50 MHz. Stations at larger distances need a kernel diameter that is proportional to the square of the distance from the centre of the array. Including the European stations of LOFAR at typically 600 km from the core would lead to unacceptably high values of the complex convolution kernel. The kernel diameter is also proportional to the square of the FoV diameter, and suggest subdividing the total FoV in a number of smaller facet fields.

Our second analysis showed that Fast Fourier transformation (FFT) of a series of small facet fields needs about the same processing as transformation of a single large field by virtue of the logarithmic characteristic of the FFT. The important conclusion here is that processing capacity required for 2-D FFT facet imaging scales with the total number of surface resolution elements in the total FoV area that needs to be imaged irrespective of the number of facets that is required to get accurate imaging over each facet.

Our third result is obtained by relating the FoV of a facet expressed in resolution elements to the integration time and relative bandwidth of the visibility samples of

the longest interferometer. When integration time and bandwidth of a sample are doubled the diameter of the facet FoV is halved and we need four facets to cover the FoV, but each facet has then only a quarter of the number of samples. The total amount of data that needs to be imaged stays the same, but the quasi-convolution step needs a smaller complex kernel to correct for non-planarity in each 2-D FFT facet image. This analysis is the basis of the proposed Fast Faceting algorithm that produces a large number of small visibility datasets that each fringe-track their own facet centre. When such Fast Faceting is implemented on the correlation platform for LOFAR using baselines up to 1200 km it would require about the same processing capacity as is needed for correlation. However, it allows reducing the output rate of the platform by selecting only the facet datasets that are required for calibration and astronomical analysis.

Fast Faceting is in a sense the array complement of multi-beaming at station level, where convolution and FFT processing produce all possible beams in the FoV provided by station and receiver system. Although all possible beams are generated, at lower processing cost than required for a limited subset by direct Fourier transformation, only a relevant subset is chosen for transport and cross-correlated. This approach greatly reduces the initial cost of a synthesis array and allows upgrading of the FoV in a later stage [Bregman, 2010].

Our fourth result is that for synthesis observations shorter than about 10 minutes the synthesized snapshot approach is the most attractive one that will in most cases not need faceting since only intrinsic non-planarity needs to be corrected in a coordinate system with a reference plane that has minimum distances to all of its stations.

Our fifth result is that correction for rotation of polarization direction of a sky source with respect to the dipole like element antennas in a phased array station has to be applied only once every 10 minutes resulting in less than 0.01% loss in signal to noise ratio of the linear polarization. The correction for polarization rotation could be combined with correction for polarization conversion using a Mueller matrix as function of position in the station beam and could be applied in the image domain. Such image correction requires all 4 polarized coherency images even when only a single Stokes parameter needs to be imaged. However, differential Faraday rotation needs correction of the full polarized coherence matrix in the visibility domain using a Jones matrix per station and could be applied in the calibration stage. Combining with corrections as function of position in the station beam needs however a quasi-convolution operation, which operation could be performed during image forming.

Our sixth analysis of correction for the polarization over the beam of a phased array station shows that bi-scalar self-calibration on an un-polarized source already corrects for the crosstalk of total intensity I to Stokes parameter Q . This effect needs to be taken into account when the Mueller matrix is constructed that converts the 4 observed and self-calibrated polarization coherencies of each image pixel to the 4 true source Stokes parameters. This approach corrects not only for polarization

rotation, but also for the shape of the station beam and for the polarization of the average beam of the element antennas in a phased array station. When the whole station beam is corrected for polarization with a single Mueller matrix valid for a single position, residual beam polarization will grow only linearly with distance from this position and reach for the HBA antenna at most 3 % at quarter power level, and much less over a smaller facet.

Our seventh observation is that in the case of synthesized snapshot images, we could at the same time also correct for variation in the actual beam shape as induced by foreshortening of the station beam of a phased array antenna station. In addition, we could correct for blind angle effects in the average element antenna beam pattern, since these effects need only a correction once per 10 minutes.

Our eighth conclusion is that a full synthesis image which takes about 6 h observing requires a number of synthesized snapshot images of typically 10 minute duration when only intrinsic non-planarity is to be corrected, which minimizes the required number of facets. This set of synthesized snapshot images needs to be integrated in a sky image after appropriate parallactic rotation and scaling corrections for each facet. This is an equivalent process as used in 3-D imaging where sub images in a number of large 2-D planes fill a volume of which only image values are needed that are interpolated onto the surface of the spherical cap that covers the full FoV. Both approaches could in practice be affordable for continuum observing with arrays with more than 30 stations and then total processing capacity will be dominated by source subtraction if more than 10 sources have to be subtracted. Both approaches combine 2-D FFT results with an intrinsic constant point spread function (psf) for each FFT grid, but the final image where each grid is rescaled separately suffers from a position dependent psf that complicates deconvolution methods that use subtraction of a nominal psf.

Our ninth conclusion is that if conventional polyhedron imaging would be extended using a complex convolution kernel of only 7^2 pixels the image accuracy would be greatly improved by correcting for 2nd order phase corrections, leaving at most 4th order ones as function of distance from the centre of the field.

Our tenth analysis discussed the balance between processing power needed for FFT, for convolution and for source subtraction and is based on the scaling laws for required processing capacity as summarized in the introduction of this section. Line imaging of observations shorter than ~6 h with less than ~50 stations is dominated by FFT processing and since few baseline samples are involved this dominance can even be maintained when a complex convolution kernel with larger linear extent is used to reduce the number of facets proportionally for each line image. This approach could even lead to a single facet and is then called W-projection and is attractive at frequencies above 1 GHz with a limited number of stations and limited maximum baseline such as ASKAP. In practice, line images need ~6 h to get suffi-

cient sensitivity and arrays with > 50 stations such as LOFAR and SKA are limited by convolution processing that also corrects for non-planarity.

Our eleventh analysis compared the processing power required for correlation with that for imaging.

For continuum observing longer than 10 minutes with more than 30 stations, the imaging is dominated by convolution processing. In case that the imaging platform has to keep up with the correlation platform we find an interesting ratio of processing power for imaging over correlation power given by $(0.1/\nu) (B_{\max}/D)^2$ with ν in Hz that is independent of the total bandwidth. This formula assumed a complex convolution kernel with a size of 7^2 pixels. Including subtraction of only 10 sources, which needs about the same processing capacity as convolution, we need for LOFAR at least a processing platform that has 6% of the processing power as provided by the correlation platform, when we observe at 50 MHz with the compact station configuration of 32 m diameter and with maximum baselines of 120 km. However, full FoV imaging with 1200 km baselines would require a processing platform with 6 times the capacity as needed for correlation. If we need subtraction of say 100 sources to reach the thermal noise floor we need 30 times the correlation processing power. It just means that in practice only a small subset of the total number of facets can be processed in real time with available processing resources.

The twelfth conclusion is that the exact height of a station above the reference plane of a quasi-planar array is not critical if complex convolution is used to correct for it. This principle could also be used for multi-beaming in a phased array antenna station where individual elements could deviate from the nominal plane. A convolution process in the signal domain corrects for non-planarity and provides samples on a rectangular grid for the 2-D FFT. Although all possible beams are made, only a subset is streamed to the limited set of cross-correlation platforms.

A thirteenth observation is that current processing implementations for correlation, source subtraction and convolution are indeed dominated by their interferometer based kernel activity. Overhead related to station calculation can be ignored for the typical LOFAR application with more than 40 stations and more than 50,000 spectral channels in each of the 4 polarization channels. This is *a fortiori* true for SKA.

A final important result is that each of the four discussed imaging approaches could be attractive for a specific application where required processing capacity has to be balanced against the inconvenience of too small facet images and complicated deconvolution procedures. Although all approaches need to give the same results for the final images within the FoV, artefacts by objects outside the FoV for which the used approximations are no longer valid might be different. This is especially true for non-linear operations such as rotations and position dependent scaling. Nevertheless, first order estimation of these artefacts can most easily be described in a coordinate system that is best adapted to that situation. Especially the 2-D FT in a coordinate system with a reference plane that has low deviations for the sta-

tions in a quasi-planar array is then the simplest reference model for analysing artefacts related to changing elevation in a long sky synthesis image.

A long synthesis observation is subsequently built up out of a series of synthesized snapshot images where aspects of sky rotation and projection can easily be interpreted as a combination of hemispheric images rotated around the polar axis.

In particular, this means that the station beam of a synthesis image is just the average of the station beams of all synthesized snapshot images. Instead of correcting for individual beams, we could simply calculate the average station beam pattern just as the average synthesized beam pattern. In principle, a weighting per synthesized snapshot could be used that optimizes the signal to noise ratio over the synthesis image. Of course, we need proper correction for average Faraday polarization rotation and instrumental polarization rotation and polarization conversion per synthesized snapshot; otherwise, the polarized signal in the synthesis observation could be reduced relative to the un-polarized part.

Recommendations

The main results and conclusions can be summarized as recommendations for implementation:

- Fast faceting technique on the correlation platform,
 - Select relevant subfields to reduce output data rate,
 - Particularly attractive for wide field continuum observing with baselines longer than 100 km.
- Facet imaging with minimum size complex kernel for quasi-convolution,
 - Minimizes convolution processing and gives facets that suffer only from 4th order phase errors for baseline projection perpendicular to image plane.
- Synthesized snapshot imaging for arrays with stations up to 100 km from the central core,
 - Slanted baseline projection along field direction on plane of the array gives 3rd order phase terms that limit duration of a tracked field,
 - Minimum size complex quasi-convolution kernel corrects mainly for 2nd order phase terms,
 - Allows a large contiguous field spanning a full station beam.

Final conclusions and recommendations can only be made when the effects of self-calibration are reviewed in chapter 4 and the effects of source subtraction, U,V-distribution and simple deconvolution on the final noise floor have been analysed in chapter 5.

4 Ionosphere Pathlength Variation and Self-Calibratability

At low frequencies, the ionosphere produces the dominant phase disturbances in the wavefronts of the signals from celestial sources [Taylor, 1999]. A monochromatic interferometer senses the phase difference induced by delay disturbance between two positions in the wavefront after passing the ionosphere. These phase disturbances are related to differences in the average refractive index along the ray paths at the two respective positions in the propagating wavefront.

As long as we are in the refractive regime we can attribute the observed interferometer phase difference to a difference in excess pathlength between the locations where rays of a source traverse the ionosphere almost parallel towards the two stations that form the interferometer. In contrast, in the diffractive regime rays are no longer parallel and intensity variation could occur between stations and even over the aperture of a station. In such situations calibration approaches that assume a refractive regime are no longer possible. The difference of the refractive index from unity integrated over the total geometric pathlength is the so-called excess pathlength and is determined by the column density of free electrons in the ionosphere also-called Total Electron Content or TEC. In practice the refractive index varies along each ray leading to ray bending that depends on the zenith angle of the incident ray [Thompson, 2004].

The basis for high quality wide-field imaging is the use of self-calibration where a number of sources in the observed field are used to determine the actual complex gain at each location in the beam of each station. With these complex gain factors the strongest objects in a field are subtracted accurately from the observed visibility data and the remaining visibility data can then be transformed to an image. To avoid artefacts from strong sources outside the field of interest we need also to subtract all those sources with the appropriate complex gain factor for each station. The conventional imaging approach using a convolution process and a 2-D Fast Fourier transformation has limited accuracy as discussed in chapter 3 and creates error side lobes in addition to the nominal side lobes of a point source. The nominal side lobes could be removed from the image by a deconvolution process that uses successive subtractions of the nominal point spread function (psf), but the error lobes remain and limit the final noise level in an image.

Self-calibration at high frequencies assumes that the complex gain factor determined from the strongest source in the field is adequate to calibrate the whole field accurately. Unfortunately, at low frequencies, where LOFAR operates, station beams are much larger and the ionosphere induces strong phase variations over the beam that are different for each station. Moreover, the station beam of the

phased array stations changes shape gradually as discussed in chapter 3, which can however be predicted accurately and source amplitudes could be corrected accordingly. More serious is that ionosphere induced phase changes vary at time scales of order $10 - 10^3$ s and only a limited number of sources is strong enough to allow every 10 s a proper complex gain estimation at every station beam given available bandwidth and station sensitivity.

It has therefore been proposed to extend single source self-calibration by using a phase screen model [Noordam, 2000], [Cotton, 2004], [Lonsdale, 2005]. Such a model needs at least the 5 strongest sources within the station beam and then allows for each telescope beam a complex gain description with one offset, two tilt and two curvature parameters. In contrast with previous generation low frequency instruments that suffer from narrow bandwidth and low aperture efficiency, this number of sources can just be obtained by LOFAR. This allows deriving a screen with delays described by a second order polynomial for each station. That is, the size of the LOFAR stations has specifically been chosen such that this second order polynomial gives a reasonable description of the phase disturbance by a TID over the actual beam size. It will be shown that the accuracy with which a delay screen over each telescope can be estimated is not only determined by the Signal to Noise Ratio (SNR) of the weakest reference source in the station beam but, more importantly, by the differences between delay values from a simple interpolation model and values according to a physical evolution model. First-order estimates will be given for these differences described by physical processes such as TID propagation and Kolmogorov turbulence that ultimately define the accuracy with which sources can be subtracted from the data while residual signals create additional noise in the final synthesis images as will be discussed in chapter 5.

It is the purpose of the remaining discussion in this chapter to analyse in detail the magnitude of the remaining phase errors and their characteristic time scales, while chapter 5 will discuss the impact of these phase errors on the effective noise in a synthesis image.

Section 4.1 summarizes some refraction basics and defines the elements that will be used in a simplified ionosphere model that describes refraction by large, medium, and small-scale structure.

Section 4.2 presents a simplified atmospheric model and compares refraction by troposphere and ionosphere as reference for contributions by a phase delay screen over the stations of a synthesis array. The model used in the discussion is characterised by curved slabs where the tropospheric ones have constant height and thickness and a homogeneous index of refraction. The ionosphere model uses three slabs where a thick curved one with constant height, thickness and homogeneous index of refraction over the array provides the TEC of order 20 TECU, but the parameters vary during the day. This change is to first-order described by a thin curved wedge on top of the thick layer that has the same refractive index. Finally

there is a thin layer at the bottom where the Travelling Ionospheric Disturbances (TIDs) and turbulence induce TEC variation of order 0.2 TECU.

Section 4.3 introduces the medium scale TIDs that define the structure of a phase screen over the beam of individual stations that form a synthesis array, together with Kolmogorov turbulence that defines the small-scale structure. Theoretical models are compared with published results of GPS and interferometer measurements to derive characteristic values for the TEC structures and their time variability.

Section 4.4 summarizes the proposed multi-source self-calibration approach adopted for LOFAR and its basic requirements, pointing out two essential issues. One issue addresses the practical problem of finding sufficient sources with sufficient signal to noise ratio to span a phase screen. The other issue addresses a fundamental problem in deriving a delay screen from observed interferometer phases.

Section 4.5 derives an integrated source density relation from published differential source counts covering the frequency range 38 MHz to 1.4 GHz and using a flux dependent spectral index. An important result is the derivation of a relation that defines the fraction of sources smaller than 0.5" as function of the flux range that allow self-calibration of the European LOFAR stations using long baselines towards the Dutch stations.

Section 4.6 determines the number of expected sources per station beam that allow self-calibration for the various station combinations of LOFAR as function of three relevant operating frequencies.

Section 4.7 presents a method that allows deriving a TEC screen over a station beam from observed interferometer phase as function of frequency for at least 5 directions per station beam. The method uses a renormalization procedure for the derived station TEC values that have an arbitrary offset per direction that is the same for all stations. It will be shown that a relative frequency range of ~20% could provide sufficient suppression of contributions of sources that are not solved for to give sufficient accurate solutions for up to ~10 source directions per beam.

Section 4.8 analyses 2nd order Lagrange interpolation over a station beam between the phases at piercing points, i.e. where the rays from telescope to source intersect the delay screen, and estimates the expected interpolation errors.

Section 4.9 summarizes results presented in the conclusions of each section.

Section 4.10 presents the main conclusion that a maximum station beam width is required that defines sufficient sampling of the large-scale TEC screen over each station induced by TIDs, while the phase error in interpolated screen data is deter-

mined by a small-scale Kolmogorov turbulence contribution that is proportional to wavelength and to beam size.

4.1 Refraction Basics

This section presents an overview of the various ionosphere effects and refraction principles as input for a simple atmosphere model that will be presented in section 4.2. Subsection 4.1.1 discusses the refractive index of a plasma as function of frequency such as the ionosphere and subsection 4.1.2 relates the difference in refraction for left and right polarized radiation to Faraday rotation induced by a magnetic field in the plasma. Although other summaries can be found in the literature, for instance [Thompson, 2004], it is practical to have relevant figures at hand based on first principles that avoid confusion when different principles need to be combined in a more comprehensive model.

Subsections 4.1.3 and 4.1.4 discuss refraction by a surface and by a homogeneous wedge respectively and points out how differences in pathlength cause a delay difference that can be observed by an interferometer.

Subsections 4.1.5 and 4.1.6 show the refraction effects of a curved homogeneous slab for a single and for a dual layer model, respectively.

All presented sub models give only first-order description of ionosphere refraction effects but form together the basis for a model to be presented in section 4.2. This model allows consistent integration of refraction by a large-scale wedge and refraction by a thin delay screen that contains medium scale TIDs and small-scale Kolmogorov turbulence and will be discussed in section 4.3.

4.1.1 Refractive index of a plasma

The refractive index of an ionized medium depends on the electron density, on the frequency, on the polarization state and on the magnetic field component in the direction of propagation. When a magnetic field is present the propagation speed is different for left circular and right circular polarization and causes rotation of the polarization angle of linear polarized radiation while passing through the medium.

Following chapter 13.3 in [Thompson, 2004] we find for a circularly polarized plane wave propagating in a medium with electron density N_e along a magnetic field B a refractive index n as function of frequency ν given by

$$n = (1 - \nu_p^2 \nu^{-2} (1 - \nu_B \nu^{-1}))^{-1/2} \quad \nu > \nu_p \quad (4.1)$$

with plasma frequency ν_p

$$\nu_p = e (4 \pi^2 \epsilon_0 m)^{-1/2} N_e^{1/2} \sim (80.6 N_e)^{1/2} \quad [\text{Hz}] \quad (4.1a)$$

and cyclotron frequency ν_B given by

$$\nu_B = e (2 \pi m)^{-1} B \quad [\text{Hz}] \quad (4.1b)$$

with electron charge e , electron mass m and vacuum permittivity ϵ_0 . For an Earth magnetic field at ionosphere height we have $B \sim 5 \cdot 10^{-5} \text{ T}$ and find $\nu_B \sim 1.4 \text{ MHz}$.

The electron density N_e in the ionosphere varies between $0.3 \cdot 10^{12}$ at night and $2 \cdot 10^{12}$ at day time in a period of sunspot maximum leading to ν_p between 5 - 13 MHz and we get total reflection for waves at frequencies below the plasma frequency.

First-order expansion of (4.1) gives $\Delta n = n - 1 \sim \frac{1}{2} \nu_p^{-2} \nu^{-2}$ and after integration of Δn over a path through the ionosphere we find an ionosphere excess phase delay τ_p^{iono} given by

$$\tau_p^{\text{iono}} = - 1.3445 \cdot 10^9 N_{\text{TEC}} \nu^{-2} \quad [\text{s}] \quad (4.2)$$

where the electron column density N_{TEC} along a path between two points is expressed in TEC units (TECU) of 10^{16} electrons m^{-2} . TEC is strongly affected by solar activity and varies in North West Europe by more than a factor two between night & day, between winter & summer and between sunspot minimum & maximum [Spoelstra, 1996]. Night time values could be as low as ~ 5 TECU and a typical day time value is ~ 20 TECU. At other locations where the SKA might be located figures will be different and the results in the dissertation could be adapted accordingly.

Multiplying the phase delay with the angular frequency gives the phase retardation φ^{iono} given by

$$\varphi^{\text{iono}} = - 2.689 \cdot 10^9 \pi N_{\text{TEC}} \nu^{-1} \quad [\text{rad}] \quad (4.3)$$

and taking the derivative of (4.3) with respect to angular frequency gives the ionosphere group delay

$$\tau_g^{\text{iono}} = + 1.3445 \cdot 10^9 N_{\text{TEC}} \nu^{-2} \quad [\text{s}]$$

Note that ionosphere phase delay and the group delay have about the same magnitude, but opposite sign and both have to be added to the geometric delay of the followed path. A N_{TEC} value of 1 TECU gives at 100 MHz an excess delay of 134 ns or 40.3 m corresponding to 84.5 rad in phase.

4.1.2 Faraday rotation

In case LOFAR observes at frequencies close to the plasma frequency, we need a higher order approximation than (4.2) for the refractive index given by (4.1) that also includes the effects of the magnetic field direction and the non-uniformity of the electron density along the path. A second order Taylor series expansion of the square root in (4.2) can be derived for a circularly polarized wave propagating with an angle θ_B relative to the magnetic field direction and gives for the excess delay along the path

$$L_e^{\text{iono}} = -403 \cdot 10^3 \cdot \nu^{-2} N_{\text{TEC}} (1 + \frac{1}{4} \nu_p^2 \nu^{-2} + \nu_B \nu^{-1} \cos \theta_B) \quad [\text{m}] \quad (4.4)$$

The frequencies in this equation are in MHz and for $\nu > 1.4 \nu_p$ and this approximation is accurate to 1%. The maximum plasma frequency ν_p^{max} , below which total reflection occurs, can be determined directly, for instance with ionosonde observation. Such a reflection measurement defines the maximum volume electron density N_{max} somewhere along the path. Also the total excess pathlength can be observed directly, for instance using dual frequency GPS observation, and defines the column density N_{TEC} . Including the second order term in the expansion of the square root to arrive at (4.4) requires path integration over N_e and over N_e^2 and we assumed a constant density. Assuming a parabolic distribution of N_e over the thickness h of the ionosphere layer with a maximum value N_{max} leads to $N_{\text{TEC}} = 2/3 h N_{\text{max}}$ for the leading factor in (4.4) and will also change the factor $1/4$ in (4.4). From observed excess pathlength and maximum plasma frequency we can derive a different scale parameter h_p for a parabolic distribution, which is a factor 1.5 larger than h for a uniform slab to give the same N_{TEC} , and is of order 200 km.

An important constraint for (4.4) is that angle θ_B satisfies $\sin \theta_B \tan \theta_B < 2 (\nu^2 - \nu_p^2) / \nu \nu_B$ according to (3.113) given by [Thompson, 2004], which can for θ_B close to 90° be approximated by

$$\nu > \nu_p + 4 \nu_B / \cos \theta_B \quad (4.4a)$$

Radiation with opposite circular polarization has the opposite sign in the ν_B term and leads to Faraday rotation angle χ for a linearly polarized wave. It is convenient to express the Faraday rotation angle χ as a fraction of the excess phase retardation ϕ^{iono} for which we find

$$\chi = \phi^{\text{iono}} \nu_B \nu^{-1} \cos \theta_B \quad [\text{rad}] \quad (4.5)$$

At 100 MHz we get for 1 TECU at most 1.18 rad Faraday rotation, but at 35 MHz this is already 9.6 rad.

Important to realize is that Faraday rotation is not only proportional to TEC (by φ^{iono}) but also by magnetic field strength (by v_B) and also by angle θ_B . Especially for an interferometer where the stations observe through a different part of the ionosphere we have to take the differences between the three factors into account.

4.1.3 Refraction by a horizontal surface observed by tilted telescope and horizontal array

This subsection discusses refraction at a surface and points out how refraction of rays at a surface according to Snell's law relates to differences in pathlength that can be observed with an interferometer.

In figure 4.1 we analyse refraction of a plane wave in vacuum incident at the surface of a homogeneous planar slab with index of refraction $n = (1 + \Delta n)$ for phase propagation.

Two parallel incident rays are drawn that make an angle $\theta + \delta\theta$ with the normal on the surface. The rays continue parallel but with a smaller angle θ . According to Snell's law $\delta\theta$ follows from

$$\sin(\theta + \delta\theta) = (1 + \Delta n) \sin(\theta) \quad (4.6)$$

Note that θ refers to the angle in the slab and evaluation using the small angle approximation for $\sin(\delta\theta) \sim \delta\theta$ and for $\cos(\delta\theta) \sim 1$ we get

$$\delta\theta = \Delta n \tan(\theta) \quad (4.7)$$

In a stratified planar atmosphere the index of refraction varies from a value $(1 + \Delta n)$ at the telescope to 1 where the rays enter.

As follows from the left picture in figure 4.1 upper and lower rays traverse equal geometric length in the same stratified medium down till the upper rim of the telescope.

The upper ray has an additional geometric pathlength P outside the atmosphere that is compensated at the level of the telescope by an equal electrical pathlength P . The tilted dish telescope needs a tilt correction to create a smaller geometric pathlength $P/(1 + \Delta n)$ in the troposphere.

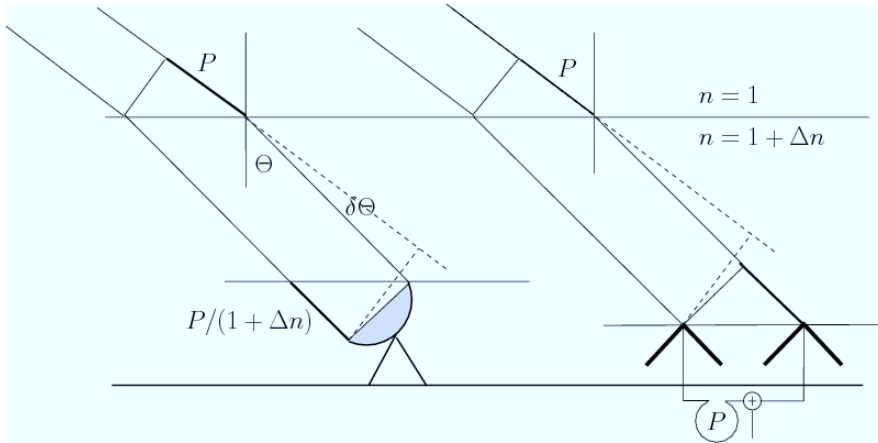


figure 4.1. Vacuum pathlength differences P in a wavefront that is refracted by a planar slab. Explanation see text.

The right picture in figure 4.1 shows the situation for a planar phased array telescope. We could assume the array in vacuum and a homogeneous planar slab gives the same exit angle above the array as at entrance. Indeed, no refraction correction is needed and a phased array antenna station is electronically steered by applying electrical pathlength corrections to the individual antenna elements that are calculated for the orientation of the baselines between the elements with respect to the direction of propagation of the incoming radiation before the atmosphere is entered.

Although the picture assumes straight rays the reasoning also holds for rays that are curved due to the stratification. For a stratified atmosphere we get on exit just above the array an additional tilt defined by the refractive index at the exit surface that is indeed equal to the bending by stratification. In a real atmosphere that is not only stratified but also curved a more complicated formula results for which (4.7) is still the dominating first-order term. An additional term, the so-called spherical refraction will be discussed in subsection 4.1.6.

4.1.4 Refraction by a wedge derived from pathlength differences

This subsection discusses refraction by a thin homogeneous wedge with small top angle, which works as a conventional prism as depicted in figure 4.2.

We assume for the wedge a refractive index close to unity, which means that refraction at the surfaces can be ignored and rays continue almost in the same direction. Only the excess pathlength differences have to be taken into account to evaluate

the refraction observed by an interferometer. For a thin horizontal wedge 2nd order effects by curvature can be ignored and we assume two parallel rays traversing the wedge towards stations in a horizontal plane that form an interferometer with baseline B . The excess delay $\delta\tau_{pz}$ in zenith direction over a wedge with length B gives for rays with zenith angle θ an excess delay difference $\delta\tau_{pz} \sec(\theta)$ when $\delta\tau_{pz}$ and baseline B are expressed in the same units. For zenith angle θ we find a refraction term $\delta\theta_{grad}$ that equals the slanted excess delay over the projected baseline $B \cos(\theta)$ given by

$$\delta\theta_{grad} = (\delta\tau_{pz} / B) \sec^2(\theta) \quad [\text{rad}] \quad (4.8)$$

and is expressed as function of the excess delay gradient and angle of incidence.

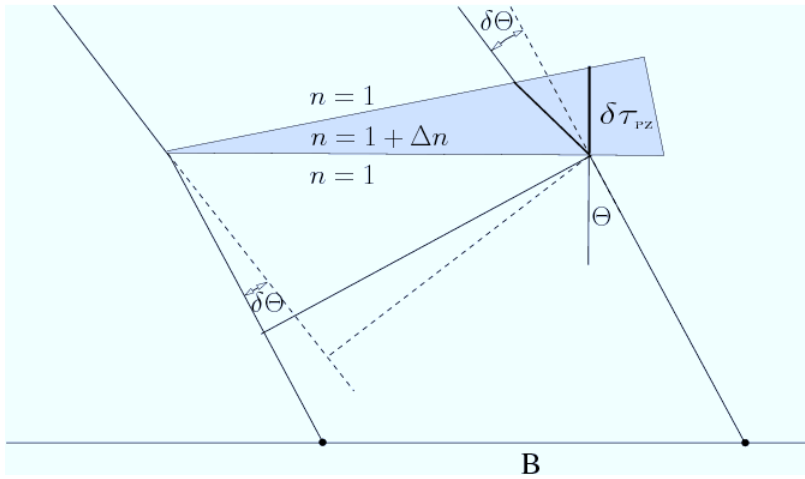


Figure 4.2. Refraction due to pathlength difference by an ionosphere wedge with negative refraction term Δn .

4.1.5 Refraction by a curved slab

A real atmosphere has not only a stratified index of refraction but is a curved slab that effectively works as a thick lens as indicated in figure 4.3.

The troposphere with refractive index larger than unity produces converging rays inside the slab. In the ionosphere with refractive index smaller than unity the rays diverge, but after passing the bottom layer the diverging rays are refracted back by a surface that is curved stronger and only a small residual convergence is left on exit as indicated in figure 4.5. As a result the ionosphere works as a positive lens and the magnification depends on the actual height of the refracting ionosphere

layer above the synthesis array. In practice we do not look along the optical axis through the centre of this lens, but along a shifted axis that could even make a large angle with the optical axis. This causes not only an image shift as by a prism, but also direction dependent field distortions.

An important effect shown by figure 4.3 is that rays with larger zenith angle follow a longer geometric path in the lower than in the upper half of the curved slab, which results in a larger contribution to the electrical excess pathlength by the refractive index in that lower half. One consequence is that a stratified refractive index leads not only to curved rays but also to a different integration of the refractive index along the geometric path.

A more detailed analysis of these effects can be found elsewhere [Thompson, 2004] and is outside the scope of this thesis. However, a simplified refraction analysis will be given in the following subsection for three reasons. In the first place it allows an estimate of the refraction caused by the geometry of a curved homogeneous slab as will be observed by an imaging interferometer.

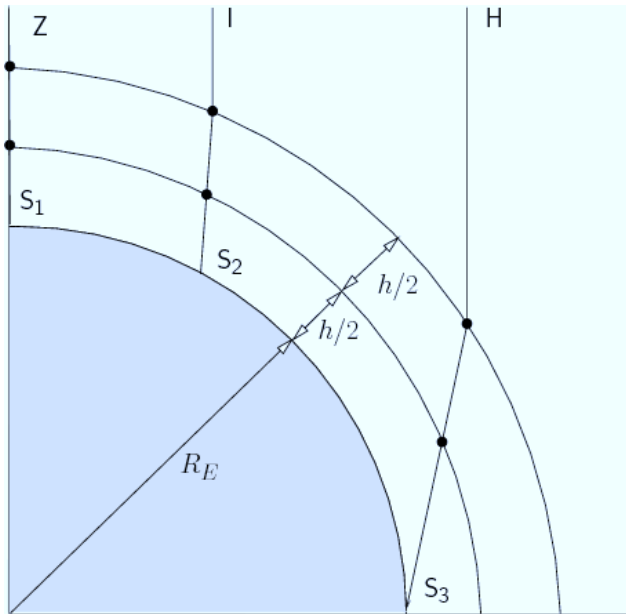


figure 4.3. Atmospheric lens showing refraction of incident rays from Zenith, Intermediate and Horizon direction to stations S_1 , S_2 and S_3 respectively. The atmospheric slab has an upper and a lower shell of equal thickness $h/2$ above an Earth surface with radius R_E .

This geometric effect dominates over other second order effects such as stratification of the refractive medium. In the second place it makes clear which assumptions are involved in deriving the final result for large-scale refraction effects and how these assumptions could impact modelling of small-scale refraction effects that use the same assumptions of low refractive index and ignoring ray bending. In the third place the resulting refraction effect can be compared with the refraction effect of the wedge component that has been discussed in section 4.1.4 and has comparable magnitude.

4.1.6 Spherical refraction by an elevated curved slab

We analyse the simplest ionosphere model with a curved homogeneous slab of thickness h at height H as depicted in figure 4.4 and compare our result with a more complicated formula that takes stratification into account to show the minor importance of the latter.

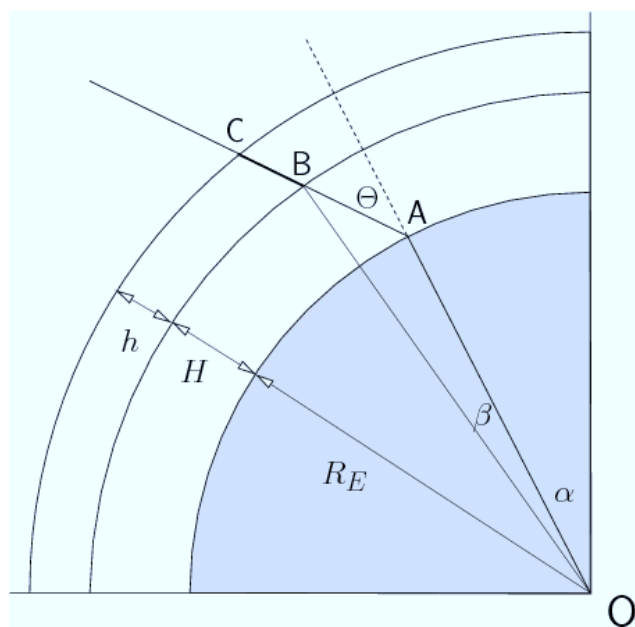


Figure 4.4. Geometry of thin slab with thickness h at elevation H above Earth surface with radius R_E . Explanation see text.

We assume for the homogeneously distributed refractive index of the slab $n \sim 1$ which implies low refraction at upper and lower surface of the slab. For a plan-parallel situation the entrance angle at the top of the slab equals the exit angle at

the bottom surface and for a small arc of the curved slab the difference between the two angles is much smaller than the refraction at either surface.

Since the latter is small, we ignore the angular refraction but concentrate on excess pathlength, just as for the wedge discussed in section 4.1.4 and we assume that ray ABC is straight. Applying the sine rule in triangle OAB gives

$$(R_E + H) \sin(\theta - \beta) = R_E \sin \theta$$

Which can be evaluates as

$$\sin \beta = (\cos \beta - (R_E / (R_E + H))) \tan \theta \quad (4.9)$$

for Earth radius $R_E = 6371$ km and $H \sim 250$ km we need $\theta < 74^\circ$ which is satisfied for LOFAR.

For small β defined by $\tan \theta \ll (R_E / 2H)^{1/2}$ we use the small angle approximation $\cos \beta \sim 1$ and we get

$$\sin \beta \sim (H / (R_E + H)) \tan \theta \quad (4.9a)$$

The excess pathlength L_e over BC is given by

$$L_e(\theta) = h (n-1) \sec(\theta - \beta) \quad (4.10)$$

Evaluation using $\cos(\theta - \beta) = \cos \theta \cos \beta - \sin \theta \sin \beta$ and using (4.9a) for $\sin \beta$ while $\cos \beta \sim 1$ gives

$$L_e(\theta) = h (n-1) \sec(\theta) (1 + (H / (R_E + H)) \tan^2 \theta)^{-1} \quad (4.10a)$$

For further analysis we use the excess pathlength L_e^Z in Zenith direction given by

$$L_e^Z = h (n - 1) \quad (4.11)$$

We consider an interferometer with two stations A and A' at angles α and $-\alpha$ with respect to the centre of baseline AA' with length $2R \sin \alpha$ and the spherical refraction angle for the projected baseline is given by

$$\delta\theta_s = (L_e(\theta + \alpha) - L_e(\theta - \alpha)) / (\cos(\theta) 2 R_E \sin(\alpha)) \quad (4.12)$$

Evaluation of (4.12) uses the derivative of $L_e(\theta + \alpha)$ with respect to α to find the increment in L_e for small α in the numerator and using $\sin \alpha = \alpha \cos \alpha$ in the denominator eliminates α while replacing $\theta + \alpha$ by θ defines the refraction for the interferometer for the zenith angle at the middle of the baseline and finally results in

$$\delta\theta_s = L_e^Z (R_E + H)^{-1} \tan(\theta) \sec^2\theta \cos^{-1}\alpha (1 + (H/(R_E + H)) \tan^2\theta)^{-2} (1 - (H/R_E) \sec^2\theta)$$

For baselines shorter than 1200 km we can ignore the $\cos^{-1}\alpha$ factor giving

$$\delta\theta_s = L_e^Z (R_E + H)^{-1} \tan(\theta) \sec^2\theta (1 + (H/(R_E + H)) \tan^2\theta)^{-2} (1 - (H/R_E) \sec^2\theta) \quad (4.12a)$$

We can compare this result with (13.137) in [Thompson, 2004], which uses ray bending in a layer of thickness h for a parabolic distribution of the electron density with peak at height H . For a parabolic distribution we have $L_e^Z = 2/3 h \Delta n_{\max} = 1/3 h (v_p/v)^2$ and assuming $h = 2 \Delta h$ we get for (13.137)

$$\delta\theta_{sp} = L_e^Z (R_E + H)^{-1} \tan(\theta) \sec^2\theta (1 + (2H/R_E) \sec^2\theta)^{-3/2} (1 + H/R_E)^2 \quad (4.12b)$$

Equations (4.12a) and (4.12b) give the same result for θ approaching zero and show that stratification and ray bending only modifies the second order correction for θ by the last two factors. First-order expansion of these terms for the converging case at $\theta = 0$ gives

$$\delta\theta_s \sim L_e^Z (R_E + H)^{-1} (1 + H/R_E)^{-1} \tan(\theta) \sec^2\theta \quad (4.12c)$$

This formula has only a few per cent error at $\theta \sim 45^\circ$ which can be compensated by adopting a different H when the functional relation is used to describe the change in $\delta\theta_s$ as function of θ over a station beam. Writing $\tan(\theta)$ as $\sin(\theta) \sec(\theta)$ shows that the factor $(1 + H/R_E)^{-1}$ just transforms $\sin(\theta)$ according to the sine rule giving

$$\delta\theta_s \sim L_e^Z (R_E + H)^{-1} \sin(\theta - \beta) \sec^3\theta \quad (4.12d)$$

This formula suggests that we need to take refraction for the Zenith angle at the place where the ray leaves the ionosphere slab at point B.

For $45^\circ < \theta < 90^\circ$ this formula becomes less accurate and even diverges at $\theta = 90^\circ$, but for practical purposes we rewrite (4.12c) for small H/R_E

$$\delta\theta_s \sim (L_e^Z / R_E) (1 - 2H/R_E) \tan(\theta) \sec^2\theta \quad (4.12e)$$

We need result (4.12e) for combination with simple geometric models using excess delay derived from TEC by integration along a straight path to allow consistent combination with other contributions that have models that also use a slanted excess delay path defined by $\sec(\theta)$.

4.2 Refraction by Troposphere and Ionosphere

In this section we introduce the basic atmospheric effects that influence imaging by interferometry such as refraction and wavefront distortion in relation to their characteristic scale sizes.

In figure 4.5 we present a simplified atmosphere model showing characteristic features of troposphere and ionosphere. The troposphere extends from sea level to about 10 km, where the lowest 3 km contains most of the water vapour that contributes significantly to the refraction of radio waves. The ionosphere extends roughly from 80 km to 600 km altitude and consists of several distinct layers (D, E, F1, F2) where F2 is the major one (thickest and highest electron density) extending from 200 km to 600 km while actual height and thickness vary following diurnal, annual and sunspot cycles. Typically the maximum electron density is at a height of 300 km, while the strongest turbulence appears at lower altitude [Thompson, 2004].

Total pathlength of rays is proportional to geometric length and to refractive index that causes delay and dispersion effects since the refractive index depends on frequency. We introduce in subsection 4.2.1 a simplified geometric model using basic elements as wedge and curved slab discussed in the previous section. For these large-scale effects we assume a constant average refractive index over a certain part of the geometrical path. For the medium and small-scale effects, we assume fluctuation in refractive index in a thin layer that change the excess pathlength for electromagnetic radiation and will be discussed in section 4.3.

More advanced refraction models combine second order effects such as stratification with curvature [Thompson, 2004], and derive higher accuracy at low elevation angles. Our combination of three simple basic models gives proper first-order estimates that allow proper assessment of the relative contributions of the different mechanisms, that each has its own interpolation method.

Subsection 4.2.2 gives an estimate for the wedge term in the ionosphere and subsections 4.2.3 and 4.2.4 give quantitative results for spherical refraction by troposphere and ionosphere respectively. Subsection 4.2.4 summarizes some conclusions.

4.2.1 Large-scale model of troposphere and ionosphere

A delay gradient as introduced by a wedge or a curved slab causes a phase proportional to baseline length, which translates to a position shift in a snapshot image. Since ionosphere excess delay is proportional to wavelength squared this also holds for delay gradients and for refraction by such gradients and both effects are proportional to path length through the ionosphere. The pathlength is proportional to the secant of the zenith angle but the projected baseline is proportional to the co-

sine of the zenith angle. As a result angular refraction by a curved slab is proportional to the tangent and to the secant squared of the zenith angle, which allows in principle separation from refraction by a gradient that is only proportional to the secant squared, as discussed in section 4.1.

We therefore separate the ionosphere delay slab in three layered sections as depicted in figure 4.5. Our simplified model has (i) a thick curved slab with a constant average zenith excess pathlength, (ii) a thin top wedge representing large-scale gradients and (iii) a thin bottom layer with small-scale gradients. The latter have short time scales and averaging over a properly selected time interval allows to separate them from the large-scale effects in top and central layer. The two large-scale effects can be separated by their different zenith angle dependence, which allows for correction of global position shift and global field distortion in a snapshot image. In principle the total excess pathlength could be determined as well and be used, together with the geomagnetic field model, for global correction of the Faraday rotation in a snapshot image. However, differential excess pathlength gives rise to differential Faraday rotation between the two stations that form an interferometer and needs to be corrected per interferometer.

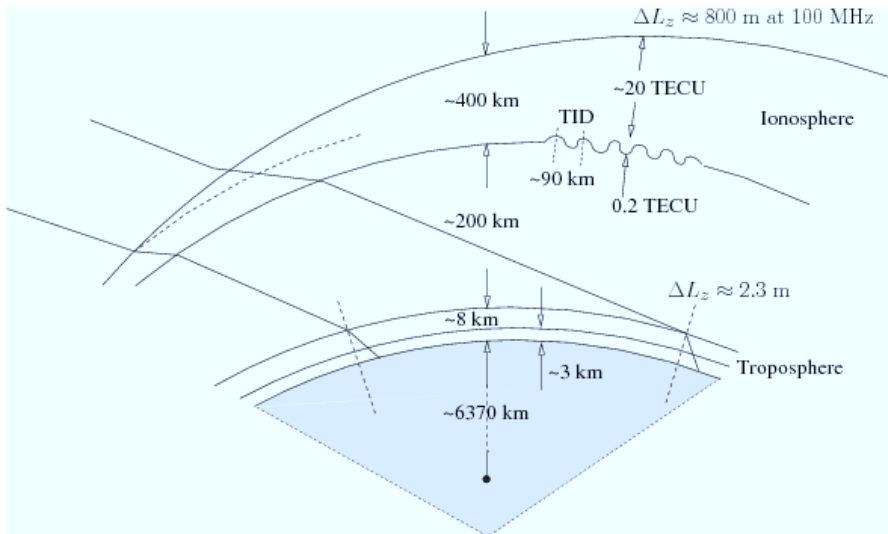


figure 4.5. Simplified tropospheric slab and ionosphere wedge model above a curved Earth surface with excess pathlength ΔL_z in Zenith direction due to refractive index. Height indications explained in text.

We make a further simplification by assuming that the layers are homogeneous. A slab of 7.8 km thickness then models the troposphere where the whole atmosphere is compressed to a homogeneous layer at standard ground pressure of 1013 hPa

and a temperature of 283 K. A plane wave entering a curved slab has different angles of incidence for stations at different locations and refracted ray paths get different electrical lengths leading to spherical refraction observable by an interferometer as discussed in subsection 4.1.6.

4.2.2 Refraction by large and medium scale wedges in the ionosphere

The large-scale ionosphere wedge has a largest thickness after local sunset and a minimum thickness just after local sunrise. At latitude 50° this corresponds to an EW gradient of the order of ~ 20 TECU over 13,000 km on top of a curved slab with average thickness of ~ 20 TECU. Over the typical array extent of order 1300 km we can therefore use a simplified ionosphere model using a curved slab with constant thickness and constant refractive index providing a column density of ~ 20 TECU in zenith direction giving spherical refraction. In addition, we have a curved geometrical wedge with the same constant refractive index but a zenith column density varying from 0 to ~ 2 TECU over 1300 km.

The basic formula (4.8) can now be used with a reasonable estimate for the TEC gradient that can be converted to a delay gradient $\delta t_p / B$ which is according to (4.2) proportional to wavelength squared. At 100 MHz we find typically

$$\delta\theta_{\text{grad}} \sim 12 \text{ sec}^2(\theta) \quad [\text{arcsec @ 100 MHz}] \quad (4.13)$$

This additional refraction is a function of azimuth, latitude and the condition of the ionosphere where the ray passes and causes a position shift.

These values should be compared with typical peak to peak variation of 0.2 TECU due to a TID with a typical wavelength of 90 km to be discussed in subsection 4.3.1. Such a medium scale wave causes a maximum gradient of 0.1 TECU over the steepest part that extends over $\sim 1/6^{\text{th}}$ of a wavelength (15 km) leading to a maximum refraction term $\delta\theta_{\text{TID}}$ that equals the slanted delay gradient over the projected baseline and equals

$$\delta\theta_{\text{TID}} \sim 53 \text{ sec}^2(\theta) \quad [\text{arcsec @ 100 MHz}] \quad (4.14)$$

This position shift varies significantly over distances of ~ 15 km, which at an assumed height of 200 km correspond to an angular extent of $\sim 4^\circ$ well within the FoV of the station beam of the LOFAR telescopes.

The resolution of an array of 12 km diameter is at 100 MHz of comparable magnitude, which means that serious blur occurs when snapshot images are averaged while a TID passes and changes sign of its refractive position shift. For larger arrays

such as LOFAR high resolution information will be averaged away in a long synthesis image if no appropriate corrections are made.

4.2.3 Spherical refraction contributions by the troposphere

Assuming that the neutral atmosphere is compressed to an air layer with uniform density at 283 K and 1013 hPa ground pressure we get a slab height $h = 7.8$ km. For refraction by the curved troposphere slab we use (4.12c) with $H = 0$ and using the refractive index for air at the quoted pressure and temperature gives $L_e^z = 2.31$ m. The spherical refraction by the troposphere is then given by

$$\delta\theta_c^{\text{trop}} = 0.07 \tan(\theta_m) \sec^2(\theta_m) \quad [\text{arcsec}] \quad (4.15)$$

For a tilted dish telescope, an additional correction according to (4.7) as discussed in subsection 4.1.3 is needed to compensate for the height difference between upper and lower rim. This is also true for an interferometer if stations have different height (above the nominal spherical Earth surface see (5.7) in [Taylor, 1999]) and we need to compensate for the additional pathlength using the average local refraction index over that height difference.

4.2.3 Spherical refraction contribution by the ionosphere

The ionosphere excess pathlength in Zenith direction can be derived from actual TEC values as for instance provided by GPS data. For a typical value of 20 TECU for LOFAR we find with (4.2) at 100 MHz $L_e^z = 806$ m and using $H \sim 250$ km we get according to (4.12c)

$$\delta\theta_s^{\text{ion}} = 25 \tan(\theta_m) \sec^2(\theta_m) \quad [\text{arcsec @ 100 MHz}] \quad (4.16)$$

This curved slab refraction by the ionosphere is proportional to wavelength squared. The presented formula has the proper leading terms with small inaccuracies starting at Zenith angles of 45° that become significant for angles $> 45^\circ$. It has a simple geometric derivation that can easily be compared with the derivations for large-scale wedge refraction and short scale TID contributions where the same simplifying assumptions are made.

Putting these refraction effects in perspective we need to realize that at 100 MHz an interferometer is needed with baselines up to 25 km to obtain a resolution of $25''$. Since spherical refraction is only a weak function of the actual height H of the ionosphere modelled as a homogeneous curved slab the actual excess pathlength L_e could be derived from the change in refracted source positions as function of observing frequency. This realization opens up a self-calibration approach for position

dependent refraction also suitable for Faraday rotation, using the model for the Earth magnetic field.

4.2.4 Summary and conclusions for large-scale self-calibration

After an overview of refraction aspects that influence a synthesis image, a first-order analysis of refraction by a curved slab was presented in section 4.1. Analysis of the excess pathlength differences between the two arms of an interferometer revealed the relative importance of geometric and physical aspects and the limitations of a homogeneous slab model. A distinction has been made between large-scale effects such as refraction based on global ionosphere behaviour that give a global shift and global distortion of an image and medium scale effects, such as TIDs, that give differential position shifts within the field of a station beam.

We described the various effects with the simplest first-order delay screen models that follow from first principles and estimated the magnitude of their effects in relation to each other. This approach allows not only making first-order estimates but also allows to identify if second order effects in a large-scale effect could influence the estimation of small-scale effects and vice versa.

Summary of results

- Analysis of tropospheric refraction revealed that the phase steering of a horizontal phased array station is not affected by refraction of a flat tropospheric slab. This is in contrast with mechanical steering of a dish telescope where the collection aperture is tilted with respect to the local horizontal plane.
- The simplest refraction model for the ionosphere is an elevated curved slab with uniform distribution of TEC giving spherical refraction, with a small wedge on top and TID waves at the bottom.
- Refraction by wedge and TID is proportional to the secant squared of the zenith angle.
- The formula for the simple elevated slab model can be simplified and shows spherical refraction proportional to the secant squared of the zenith angle and proportional to the tangent of the zenith angle.
- Comparison with a spherical refraction model for the ionosphere that includes the effect of stratification and large zenith angle shows that our simplified model is correct for zenith angle $\theta < 45^\circ$ and is adequate to describe spherical refraction by $\tan(\theta)\sec^2(\theta)$ over the extent of a station beam for $45^\circ < \theta < 75^\circ$.

Our simplified analysis shows the following well-known effects and scaling laws

- Array observable spherical refraction by a curved tropospheric slab is only a small fraction of the refraction as observed with a telescope where an aperture is tilted with respect to the local horizontal plane.
- Medium scale effects such as TIDs can cause local position shifts in an image of even larger magnitude than large-scale refraction. These local shifts vary over angular distances of order 2° on the sky and are proportional to the squared secant of the zenith angle.
- The large-scale thin wedge causes a position shift that is about a factor three smaller but is also proportional to the square of the secant of the zenith angle.
- All refraction is proportional to the TEC along the path through the ionosphere and therefore proportional to wavelength squared which allows easy separation from effects with a different scaling law such as tropospheric refraction (which is at 100 MHz a factor 350 smaller) and clock errors which are independent of wavelength.

Conclusions for large-scale self-calibration

- Absolute refraction can be determined for each ~ 10 min image for the field as a whole when sufficient resolution is present as provided by baselines longer than ~ 15 km. In addition, we need a relative bandwidth of order 20% that allows solving for a quadratic increase by 44% of this refraction with wavelength.
- Since the curved slab has an additional $\tan(\theta)$ dependence in its refraction contribution, it can in principle be separated from the thin wedge contribution in a longer observation. When this could be realized effectively, the TEC of the uniform slab alone can be obtained.
- This observed TEC could then be used for Faraday rotation correction of observed polarization angles over the imaged field with the aid of the geomagnetic field model.
- Differential Faraday rotation between the two ray paths of an interferometer is proportional to differences in excess pathlength as caused by refraction such as by the thin wedge, by TIDs and by a curved slab and needs a separate polarization rotation correction for each station.

4.3 Ionosphere phase delay screen contributions

Our focus in this section is on medium and small-scale structure in the wavefronts such as caused by TID and Kolmogorov Turbulence over the stations. The induced phase disturbances limit image quality and even the effective noise after self-calibration on objects within the field defined by the station beams. We need however to realize that these structures are embedded in large-scale wavefront deformations by spherical refraction and wedges that determine the global position of the observed field relative to the sky. It is therefore important to know the mechanisms of the various contributions to the observed phases to invoke adequate correction procedures for each disturbing mechanism using its characteristic time scale and functional dependencies.

The phase screen used in calibration models is in fact a thin slab with varying pathlength as function of location while the phase is the product of total pathlength and frequency. The pathlength is proportional to local thickness, to the secant of the inclination angle and to the refractive index, which is a function of frequency itself. Variation in phase as function of location has therefore many origins and we simplify the phase screen model by assuming a thin delay screen with constant thickness while all delay variation is attributed to variation in effective refractive index.

The phase corrections for small spatial scales derived from this simple screen model need to be combined with phase corrections for refraction derived from models that describe large spatial scales. We want to assign observed phase variations at a few locations either to a screen contribution or to refraction to allow proper interpolation to even smaller spatial scales. To this end, we extend the thin delay screen with varying refractive index by a thin wedge that has constant refractive index. A curved slab of constant thickness and constant refractive index describes the bulk ionosphere. This curvature causes so-called spherical refraction that causes image distortion by systematic position dependent shifts. The thin wedge of excess pathlength causes an additional shift of the whole field that varies on a diurnal scale.

Kolmogorov turbulence (KT) and Traveling Ionospheric Disturbances (TID) [Thompson, 2004] define fine structure at timescales between ~ 10 and $\sim 10^3$ s respectively, which lead to random instantaneous position shift and image blur that are position dependent. The main effect is a reduced peak flux for point sources in a synthesis image if data is averaged over long timescales without short term corrections.

Although phase is the observable by a narrow band interferometer, LOFAR and SKA have sufficient bandwidth to derive delay parameters that can be attributed to an origin in the ionosphere such as the total electron content along the path of the wavefront towards the telescopes.

Subsection 4.3.1 introduces the TIDs and subsection 4.3.2 summarizes the observable results of refraction by Kolmogorov turbulence. Subsection 4.3.3 compares the predicted results with observed interferometer data showing consistent results. Subsection 4.3.4 summarizes the conclusions.

4.3.1 TID waves in lower ionosphere

A layer with fluctuations in refractive index distorts a propagating plane wave and causes phase differences over the plane of the wave. These phase differences can be observed with an interferometer and increase with increasing separation between two positions in the wavefront up to the maximum scale size of the fluctuations.

The most important small-scale phenomena are the so-called Travelling Ionospheric Disturbances (TIDs), which are a manifestation of acoustic-gravity waves in the lower ionosphere. The associated fluctuations in electron density produce a wave pattern in the differential excess delay as observed with an interferometer and the amplitude of the delay pattern scales quadratic with wavelength.

There are three distinct categories of TIDs [Velthoven, 1990]:

- LSTID: large-scale TIDs have a horizontal phase velocity substantially larger than the MSTIDs and SSTIDs (Medium and Small-scale TIDs), namely 300-1000 m/s. Periods range from 30 min to 3 hours and wavelength exceeds 1000 km. Propagation is equator-ward from polar regions, where they are supposed to be generated in the auroral zones. The mechanisms that generate LSTIDs is a topic of ongoing scientific research.
- MSTID: horizontal phase speeds of 100-300 m/s, period from ~12 min to ~1 hour, wavelength of several hundreds of km. Occur much more frequently than LSTIDs. Origin is unknown but several candidates are proposed (orographic effects (mountains, etc), wind shears, the terminator, tropospheric effects, atmospheric tides, etc.).
- SSTID: periods of several minutes and wavelengths of tens of km. Associated with acoustic branch of the AGW (Acoustic-Gravity Wave) spectrum. Origin is unknown.

The most relevant Medium-scale TIDs have quasi-periods of 10-20 min and scale lengths of 100-200 km with 0.5-5% variation in TEC [Thompson, 2004]: "The atmosphere has a natural buoyancy, so that a parcel of gas displaced vertically and released will oscillate. Shorter wavelengths correspond at an assumed propagation speed of ~150 m/s to shorter periods for which pressure is the restoring force. Longer wavelengths correspond to longer periods for which gravity is the restoring force". At an assumed height of 200-250 km, a short wavelength of 90 km converts to an extent ~24° on the sky and a single sine like wave pattern can then be approx-

imated by pieces of $1/6^{\text{th}}$ of a wavelength or 4° on the sky. Such a piece of a sine wave pattern represent either a linear delay slope for the interval $[-\pi/6, \pi/6]$ or a parabolic one for the interval $[2\pi/6, 4\pi/6]$ while pieces of 2° as for the interval $[\pi/6, 2\pi/6]$ have an intermediate delay shape. For an array with an extent much smaller than $1/6^{\text{th}}$ of a TID wavelength, the linear delay gradient causes for all sources in the direction of the flat part of the wave a constant position shift along the direction of propagation of the acoustic wave. A parabolic gradient gives a varying position shift within a maximum sky extent of 4° . This shows that the 2-D wavefront distortion over the beam can indeed be described by a limited set of 5 parameters, one for offset, two for tilt and two for curvature but only if the station beam is indeed narrower than 4° . Derivation of these 5 parameters needs at least 5 sources to be observed within a coherence interval each with sufficient signal to noise ratio on most baselines, which will be further analysed in section 4.2.

The parabolic part of the delay gradient over the aperture of an array gives a position shift of objects that varies with position within the station beam. In addition, the curved wavefront blurs objects in an image. This blur is small for an array with an aperture much smaller than $1/12^{\text{th}}$ of a short medium scale TID wavelength, i.e. much smaller than 7 km. Aperture diameters of 7-15 km give snapshot images where point sources suffer not only from a position dependent shift but also from serious blur by the observed curvature in the delay screen. Arrays larger than 15 km could even produce snapshot images with speckle patterns instead of point sources.

Maximum TEC gradients of 0.1 TECU over 15 km ($1/6^{\text{th}}$ of a TID wavelength) give sources within an image field a maximum position shifts of $53''$ @ 100 MHz, which is larger than the resolution of an array of 12 km diameter at that frequency. Averaging over many snapshots while such TIDs pass then results in image blur limiting the effective resolution of a synthesis image. Synthesis images by larger arrays suffer from blur by shifting of sources around an average position, but also each point source is a sum of "speckled" snapshot images. Each snapshot image needs therefore appropriate phase correction per telescope and per source direction. Such corrections need estimation of an excess pathlength difference between the delay screen common to all stations in the core and the screen at each station further away than 7 km from the centre of the array. These delay screens can then be used to correct the phases for all sources on all baselines between all stations.

The effects of large-scale TIDs can modelled by our wedge term discussed in subsection 4.2.2 requiring correction at time scales of ~ 10 min and can be combined with other corrections for that time scale such as changing spherical refraction and Faraday rotation.

4.3.2 Kolmogorov turbulence model

TIDs create refractive index fluctuations up to 0.2 TECU at horizontal scales larger than 50 km [Thompson, 2004]. These large fluctuations break down into smaller ones at shorter distance scales with accordingly shorter time scales, as follows from their quasi-periodic behaviour with fine structure as shown by figures 4.6 and 4.7. It is therefore reasonable to expect a further breakdown into smaller structures and analyse the effects on wavefront propagation.

V.I. Tatarski and D.L. Fried derived between 1961 and 1971 useful expressions for optical propagation through a turbulent troposphere using statistical models. An important concept is the assumption of a “frozen” distribution of irregularities that “dissipates” only slowly, while the screen movement in horizontal direction causes more rapid temporal fluctuation in the pathlength of the rays that cross the screen in an astronomical observation [Thompson, 2004], [Tol, 2009].

Tracking sky objects from Earth rotating at $0.25^\circ/\text{min}$, results in a maximum relative screen movement of 18 m/s for a co-rotating frozen screen at a height of 250 km. GPS satellites move at 3.8 km/s at 20,200 km above the Earth surface, which results at a height of 250 km in a relative screen speed of 47 m/s when observed near zenith, while the medium-scale TIDs have a propagation speed of order 150 m/s. This combination of density propagation in the delay screen and decay thereof, while the whole screen moves relative to the ray path complicates the analysis.

Although the medium scale TID effects can be reasonably modelled by virtue of their origin, the small-scale Kolmogorov contributions cannot, but have decreasing effects at smaller scales and reach a level where the induced phase fluctuations are comparable to those by calibration accuracy as will be discussed in sections 4.7 and 4.8.

4.3.2.1 1-D analysis of GPS track data to define variation over phase screen

Ionosphere excess delay on the path between a GPS satellite and a receiver can be derived from a dual frequency observation. With a dedicated receiver the phase difference between signals transmitted at frequencies $L1 = 1575.42$ MHz and $L2 = 1227.6$ MHz is observed and after a number of corrections the excess delay along the path through the ionosphere can be derived.

A series of GPS observations made in The Netherlands during January 2006 have been analysed [Tol, 2009] and the excess delay converted to an excess phase at 74 MHz along the ray path. Figure 4.3(d) in [Tol, 2009] shows the rms phase difference between two positions along a satellite track, which is averaged over ~1000 tracks and plotted as function of separation s . The following model, valid for zenith direction, can approximately describe this graph

$$\sigma_{\varphi}(s) = (s/s_0)^{0.8} \quad [\text{rad}] \quad (4.17)$$

Valid for $1 \text{ km} < s < 100 \text{ km}$ and $s_0 = 2.0 \text{ km}$ at 100 MHz .

Since phase scales for a given separation proportional to wavelength, the distance parameter s_0 scales to other frequencies ν as $\nu^{1.25}$, which has been used to convert from 74 MHz to 100 MHz .

The observed rms phase differences over many days actually follow a power law with exponent varying from 0.65 to 0.9 with an average value of 0.8 close to $5/6 = 0.83$ valid for a Kolmogorov turbulence model. For the latter model a number of relations have been derived using the Fried parameter r_0 which is the diameter of a coherence cell defined as $r_0 = 3.18 s_0$ [Thompson, 2004]. In this new unit we find for the rms phase difference between two points in a wavefront separated by r

$$\sigma_{\varphi}(r) = 2.62 (r/r_0)^{5/6} \quad [\text{rad}] \quad (4.18)$$

This Fried parameter is most useful in 2-D averaging over an aperture.

4.3.2.2 2-D tip-tilt correction and residual deviation over small area of phase screen

The rms phase over a circular aperture area A with diameter B can be evaluated from (4.18) and according to [Thompson, 2004] we get

$$\sigma_{\varphi A}(B) = 1.01 (B/r_0)^{5/6} \quad [\text{rad}] \quad (4.19)$$

A phase gradient over the filled aperture A of a telescope with diameter $B > r_0$ leads to an angular position shift that has an rms value according to (4.63) in [Tol, 2009] with reference to a 1965 publication of D.L. Fried and is given by

$$\sigma_{\alpha A}(B) = 0.6 (\lambda/r_0) (r_0/B)^{1/6} \quad [\text{rad}] \quad (4.20)$$

For a circular aperture with diameter B we could determine a best fit phase gradient over a wavefront that is distorted by Kolmogorov turbulence and find after subtraction of this tilted phase plane a residual phase distribution. A simplified first-order approximation for the variance of the residual phase over the aperture with diameter B after tip-tilt correction can be derived from (4.19) and (4.20). The variance of a tilted plane is found by integrating $(2\pi \sigma_{\alpha A}(x) / \lambda)^2$ over $-B/2 < x < B/2$ with aperture coordinate x in the direction of the tilt. We take the difference with the variance $\sigma_{\varphi A}(B)^2$ as given for the rms phase of the aperture (4.19) before tip-tilt correction and find as first-order estimate

$$\sigma_{\phi A}(B) \sim 0.66 (B/r_0)^{5/6} \quad [\text{rad}] \quad (4.21)$$

Equations (4.19) and (4.21) are the basis for high resolution imaging with small optical telescopes. Depending on troposphere conditions the seeing cell diameter r_0 varies between 0.1 and 0.3 m at a wavelength of $5 \cdot 10^{-7}$ m and limits the resolution of a telescope with a diameter much larger than the seeing cell diameter to $\sim \lambda/r_0$ with values between 1" and 0.33" respectively. By limiting the aperture diameter of an optical telescope to $3 r_0$ and applying a tip-tilt correction faster than the coherence time the resolution should be improved by a factor 3. However, the residual phase over the aperture causes an additional blur limiting the effective resolution improvement to a factor ~ 2.4 . A known result of a 2-D integration for the residual rms phase of a circular aperture with diameter $B = 3 r_0$ after tip-tilt correction is ~ 1 radian, which indicates that the factor 0.66 found for our first-order estimate (4.21) is too large and should be 0.4, giving

$$\sigma_{\phi A}(B) \sim 0.4 (B/r_0)^{5/6} \quad [\text{rad}] \quad (4.21a)$$

for the residual rms phase over the aperture after tip-tilt correction.

4.3.2.3 Differential angular position shift within a station beam

Differential angular position fluctuation as function of angular separation α between objects can be derived for an aperture of diameter B when observing through a screen at height H . From the phase structure function of the screen characterized by (4.18) an expression for $\sigma_\alpha(\alpha)$ can be derived that is a function of r_0 , H and B using $\alpha = r/H$. Unfortunately there is no simple analytic expression but a numerical evaluation has been made [Tol, 2009] for a symmetric aperture filled with baselines of a synthesis interferometer. Tol discovered that the rms position shift depends on the direction with respect to the source separation. Adding the variance of the two orthogonal components results in a rms position shift as function of separation, and is depicted in figure 4.11 in [Tol, 2009] for the aperture distribution of the VLA in its B-configuration. It has to be understood that the graph gives the rms of differential position shifts normalized to the differential position shift for the case that the apertures in the two directions would be completely independent [private communication].

As long as there is an overlap of the projected apertures on the phase screen we find that the shape of the function that describes the differential shift as function of the separation depends on the exponent in the power law of the rms phase difference. Curves are given for exponent values of 0.6, 0.83 and 0.95, where 0.83 corresponds to Kolmogorov turbulence. The curves show an almost linear increase of the rms value of the differential angular position fluctuation as function of angular position difference up till some knee in the graph. Then a much smaller gradient in the rms difference appears that flattens off at angles where the rays have pierce

points at the ionosphere delay screen with a separation larger than the diameter of the array aperture.

4.3.2.4 Relevant time scales

It is important to note that the derived relations for rms phase, the rms phase gradient and the rms tilt gradient in a wave front that passed a delay screen are ensemble averages, while instantaneous observations are distorted by a particular realization of the refractive index distribution. When the station beams of a synthesis array track a source field, the piercing points of rays from the different telescopes to different sources, move over the delay screen and suffer from varying excess delay paths. The screen itself changes due to propagation of density waves while the electron content differences evolve. Apparently, this evolution can be reasonably well described by a Kolmogorov turbulence model, which needs a so-called outer scale of turbulence. We assume the TID wavelength as an appropriate physical phenomenon that happens to be roughly equal to the vertical thickness of the ionosphere, which could also be an appropriate physical scale. For a given wavelength of medium scale acoustic-gravity waves in the lower ionosphere, we get a time scale defined by the half period of the TID wave that follows from the propagation speed of 150 m s^{-1} . Any locally induced small change in density will also propagate at this speed and lead to characteristic time scales for decay of TEC differences. According to (4.17), the characteristic distance over which Kolmogorov turbulence gives a phase difference that is on average 1 rad is 2 km at 100 MHz. Such a difference could propagate with a speed of 150 m/s which then corresponds to a characteristic time scale of 13 s. However, when a frozen phase pattern is assumed that is tracked at 18 m/s the characteristic time constant for 1 rad phase change is 110 s. An interferometer observes a phase difference over a given distance separation in the delay screen which has a component determined by a propagating large-scale wave, but also a small-scale component defined by the difference between two frozen-in Kolmogorov turbulence structures that appear when a sky source is tracked.

Half a wavelength for a TID and 1 rad for Kolmogorov turbulence seem somewhat arbitrarily to define time constants, but they give at least an indication. In section 4.8.4 we will derive characteristic time scales that are more appropriate to define observational integration times that can be used to derive delay screen parameters useful for self-calibration.

4.3.3 Comparison with interferometer data

In this subsection, we compare the theoretical results for TIDs and Kolmogorov turbulence with actual interferometer data and show that the TID represents the medium scale variation while Kolmogorov turbulence is an effective description for small-scale effects.

4.3.3.1 Differential angular position shift and associated source degradation

The VLA has baselines up to 30 km that sample large separations in the delay screen that are comparable to the extent of the delay screen covered by a station beam, which allows a more detailed analysis of ionosphere effects. Observations for the VLA Low-frequency Sky Survey at 74 MHz (VLSS) [Cohen, 2007] were made in late fall 2003 and early spring 2005 under benign ionosphere conditions as expected for observing just before and during the solar sunspot minimum.

The VLSS images are averages of snapshot images with 2 min observing time. Subsequent images show variation in the observed position shift between sources in each image and allow estimation of an rms value of the differential position angle as function of separation in the delay screen. For an assumed height H of the ionosphere delay screen we can convert observed angular separation α into spatial separation r at the screen according to $\alpha = r/H$.

Two datasets from the VLSS have been used to fit differential position shifts over an area of 15° diameter to a Kolmogorov model as depicted in figure 4.12 by [Tol, 2009]. Instead of the expected Kolmogorov exponent of 0.83 a best fit value of 1 was obtained for different ionosphere conditions. Accordingly, we rescale the values for s_0 linearly with frequency from 74 MHz to 100 MHz and the angular shifts inversely squared.

A quiet ionosphere gave $H = 250$ km, $s_0 = 2.91$ km and an rms differential angular position shift that linearly increases with separation r to a plateau at $11''$ for $r > 20$ km that corresponds to angular separation $\alpha > 4.6^\circ$. A disturbed ionosphere gave $H = 210$ km, $s_0 = 2.00$ km and an rms differential angular position shift increasing to $19''$ at 10 km or $\alpha = 2.3^\circ$. For larger α the differential shift slowly increases to $36''$ at $r = 50$ km corresponding to $\alpha = 11^\circ$ equal to the width of the station beam. A separation of 10 km in the ionosphere screen leads according to (4.18) for $s_0 = 2.00$ km to a rms phase difference of 3.6 rad which at 100 MHz corresponds to a delay of 1.7 m. Such a delay gradient over 10 km gives a refraction of $34''$, as observed, and a differential refraction over the angular extent of 2.3° that is less.

The ranges of the angular extent that describe the ranges of relative position differences are therefore in good agreement with the simple TID analysis presented in subsection 4.1.4. Separation of a wave of 90 km in pieces of 2° and 4° angular extent with linear and curved slopes indeed explains the results as function of separation. The relatively large refraction value of $34''$ at the transition region between 2° and 4° angular extent is an indication that the estimated s_0 is too low. This low s_0 is indeed to be expected since the best fit exponent of 1 indicates that other mechanisms than Kolmogorov turbulence dominate for which we identified the TIDs providing characteristics as observed. This suggests that when identifiable wave patterns could be removed still smaller phase differences will be observable that

could have the characteristic behaviour associated with Kolmogorov turbulence as will be discussed in subsection 4.3.2.2.

As a first-order approximation for analysis we can assume a Kolmogorov model with $s_0 \sim 1.8$ km (at 74 MHz) giving $r_0 \sim 5.7$ km and after a best fit tip-tilt correction for the VLA aperture of 10 km diameter using (4.21a) we find a residual rms phase difference σ_ϕ over this distance of 0.64 rad. Assuming this phase noise on all baselines would lead to reduction in peak flux of a point source by a factor $\exp(-\sigma_\phi^2/2) \sim 0.82$ when many snapshots are averaged. This number is indeed confirmed by comparing peak flux values of sources in an image that was calibrated using only tip-tilt correction per snapshot (with the so-called field calibration method) with peak fluxes of sources in an image of the same data where all interferometers were calibrated per snapshot using the strongest source in the field [Cotton, 2004]. The latter method increased the peak flux for the strongest source but increased the phase errors for the sources at larger distance from the reference source, resulting in even lower peak fluxes of these sources.

This suggests that separate self-calibrations should be used for more reference sources in the field, but this requires a signal to noise ratio larger than 3 on each source at each baseline within an ionosphere coherence time as will be discussed in section 4.4. The remaining phase errors do not only broaden the observed width of a point source but create additional side lobe structure that varies with the sources over the field such that the noise floor in the synthesis image is effectively raised.

4.3.3.2 *Differential phase gradients over a large aperture*

A phase gradient over the aperture of a synthesis array causes a position shift of observed objects, so looking into different directions projects different parts of the ionosphere delay screen over the array aperture that have different gradients and give consequently a direction dependent position shift.

Actual TEC variation has been observed at ~ 139 MHz with the Westerbork Synthesis Radio Telescope (WSRT) under benign ionosphere condition around end November 2007 [Bernardi, 2010]. The WSRT has only East-West baselines up to ~ 2.8 km and samples therefore a small fraction of the delay screen to determine a local gradient in East-West direction. The station beam has a width of $\sim 6^\circ$ FWHM and allows comparison of the gradient over a considerable fraction of a TID wave.

The self-calibration data presented in figure 4.6 [Bruyn, private communication] were analysed and confirm the previously quoted temporal periodicity while showing $18\text{-}130^\circ$ peak to peak phase variations over 2.7 km, which converts to rms TEC gradients with magnitude of 0.001 - 0.008 TECU/km.

Since the WSRT is an EW oriented linear array the direction of the TID waves cannot be identified, only their period in time, which allows only a first-order estimate of the actual phase slope of a TID wave. For a TID with an assumed wavelength of 90 km we derive peak to peak values of 0.08 - 0.19 TECU if we also assume EW propagation. These results are consistent with data at ~330 MHz, ~610 MHz and 1420 MHz over the period 1970-1991 [Spoelstra, 1996].

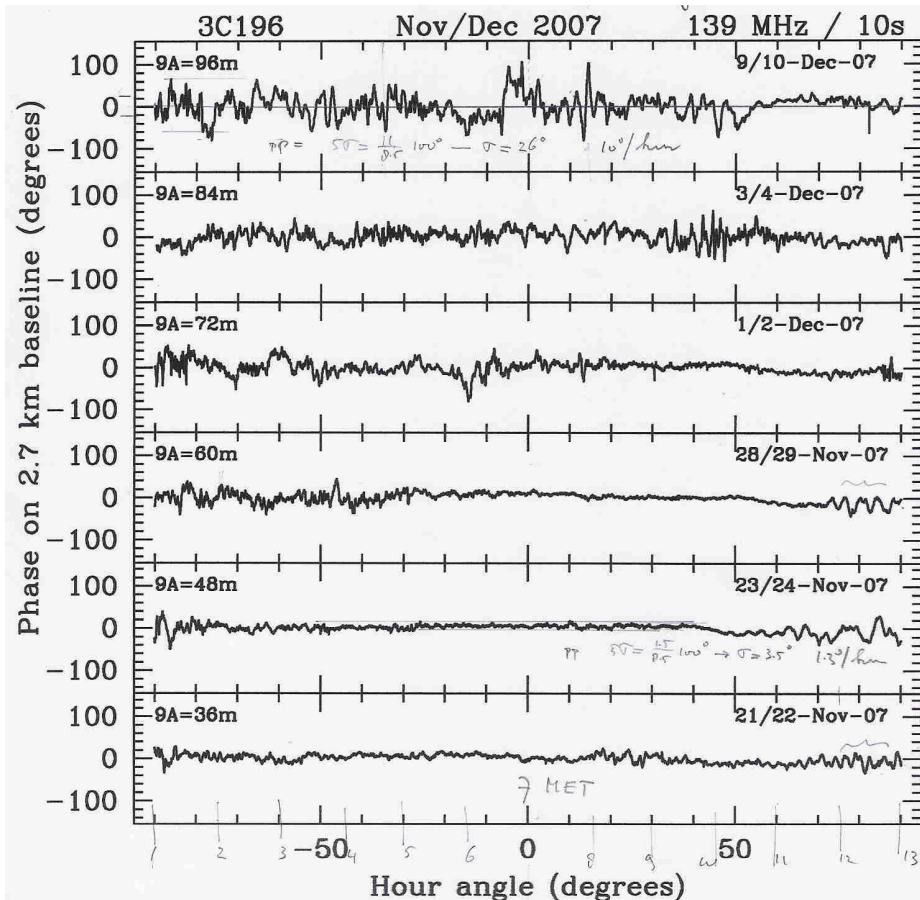


Figure 4.6. Phase difference at 139 MHz between stations separated by 2.7 km attributed to ionosphere (by courtesy of A.G. de Bruyn).

The WSRT data [Bernardi, 2010] have baselines comparable to scale length s_0 derived from GPS data for The Netherlands discussed in subsection 4.3.2.1 and we can use (4.17) to make a first-order estimate of the scale length using the peak to peak phase difference over 2.7 km from figure 4.6 for the quiet period. If we define

the rms as $1/5^{\text{th}}$ (in view of the limited number of variations) of the peak to peak value we find $\sigma_{\phi} \sim 3.5^{\circ} - 26^{\circ}$ at 139 MHz. After inverse proportional scaling of the phase to 100 MHz we find s_0 values of 58 - 4.8 km. Inspection of the WSRT synthesis images [Bernardi, 2010] shows that if the visibility data are corrected for the shift of the central source that sources at 1.8° , 2.3° and 3° separation still suffer from relative position shifts of about the same magnitude as the correction. This result is consistent with the prediction for differential shifts by the TID model as discussed in subsection 4.3.2.1. Separate self-calibration on these sources, which provided sufficient signal to noise ratio per baseline per 10 s integration time, allowed accurate removal of their artefacts from the synthesis image.

Figure 4.6 shows 6 observations in a 3 week winter period during sunspot minimum, which should give low ionosphere phase disturbances. Indeed we find one observation with extremely low phase fluctuations at night time even continuing in the early morning after sunrise. There is also one observation with very large phase fluctuations during night and morning that decay around noon. About half of the observations show clear wave patterns developing around noon with periods characteristic for medium scale TIDs. These results suggest that correcting for the phase by these waves leaves low residual phase fluctuation warranting high quality low-frequency imaging for about half of the observations in a period of benign ionosphere condition.

4.3.3.3 Large-scale TID and small-scale Kolmogorov Turbulence results

Although GPS data [Tol, 2009] support Kolmogorov turbulence as expressed by an average exponent of 0.8, the variation of the exponent over 24 hours indicates the existence of additional mechanisms such as TIDs. Our figure 4.6 clearly indicates such differences between day and night and between different days, which implies that a single s_0 of 2 km from a long term average is biased towards the lower values valid for worst case ionosphere conditions, while in practice mostly high quality observations will be selected for further processing.

A recent LOFAR observation (July 2011) using a 2-D interferometer distribution with 25 km baselines [Bemmel, daily image 20110914, www.astron.nl] shows global half-periods of 10 - 15 min with peak to peak values of 0.04 - 0.12 TECU. A reproduction is given in figure 4.7.

An interferometer samples the instantaneous spatial derivative in the delay screen, so figure 4.7 shows therefore not only propagation of a wave but also evolution as function of time. Further analysis could provide in principle estimates of the actual height of the layer as well as direction and speed of propagation of the TID. Apparently, a spectrum of periods is present creating fine structure with half periods of 5 min.

A simple graphical fit to the curves as could be provided for instance by a large-scale sine wave pattern or a second order polynomial leaves fine scale residuals extending over ~ 15 km with a typical maximum deviation of ~ 0.005 TECU from a large-scale quasi-periodic structure with assumed wavelength of 90 km. These results are consistent with LOFAR data over the 33–68 MHz range [Tol, 2011] that indicates a Kolmogorov model with large-scale parameter $s_0 \sim 50$ km at 100 MHz once the TID effect is removed. The same value is found for one of the observations by [Benardi, 2010] that had no TID. If a large-scale delay screen model corrects the visibility data for the TID, we find according to figure 4.7 residual small-scale structures of ~ 0.005 TECU. These will result in 0.3 rad rms phase noise at 140 MHz that could potentially give 5% degradation in point source flux if random for all interferometers. In fact, the correction for the TID is a tip-tilt correction leaving residual phase errors over the aperture given by (4.21a) that give for $r_0 \sim 6$ km the same residuals. A more extensive analysis of residual Kolmogorov turbulence effects will be given in subsection 4.8.3 in relation to the residual effects of a specific large-scale correction.

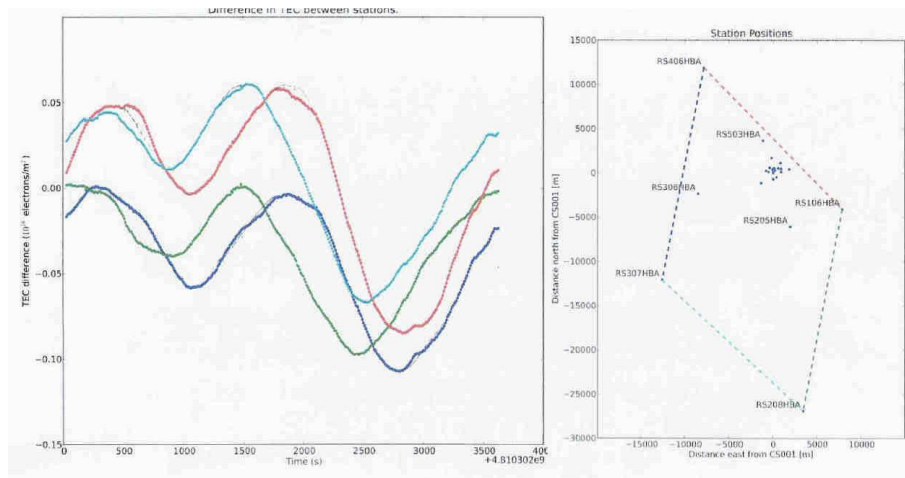


Figure 4.7. Difference in TEC between 4 LOFAR stations [Bemmel. Daily image 2011 0914, www.astron.nl].

Unfortunately, at 70 MHz we could get 0.6 rad rms phase noise and 20% degradation becoming even worse at lower frequencies. It means that a large-scale delay screen model by sampling the TID every 11 km or every 3° on the sky might not be sufficient at these lower frequencies and that modelling on finer scales is required. In practice lowest frequency observations will be done in good ionosphere conditions and only those observations will be processed that that suffer from TIDs with

low TEC amplitudes and sufficiently long wavelengths that allow proper delay screen modelling over the station beam with order 5 self-calibration sources.

4.3.4 Summary and conclusions for small-scale self-calibration

Analysis of the excess pathlength differences between the two elements of an interferometer revealed the relative importance of geometric and physical aspects. A distinction has been made between large-scale effects such as refraction based on global ionosphere behaviour that give a global shift and global distortion of an image as discussed in section 4.2 and small-scale effects, such as TIDs, that give distortions and differential position shifts discussed in this section 4.3. Although medium scale TIDs induce large phase variation over spatial intervals of 50 km and temporal intervals of 10^3 s, integration over 10^{-10^2} s intervals leads to only limited sensitivity degradation. Appreciable ionosphere effects appearing at time scales < 10 s such as amplitude scintillation are a sign of a highly disturbed ionosphere that is not suitable for high quality imaging.

We described the various effects with the simplest first-order delay screen models that follow from first principles and estimated the magnitude of their effects in relation to each other. This approach allows not only making first-order estimates but also allows to identify if second order effects in a large-scale effect could influence the estimation of small-scale effects and vice versa.

Our analysis of the literature and of available observational material can be summarized:

- Observed differential position shifts within a field of view of 10° reach a maximum for angular separations between 2° and 4° , consistent with TEC structures described by a wavelength of order 90 km and amplitudes of 0.02 - 0.1 TECU that vary at scales of half a period.
- The medium scale TID is a recognized physical phenomenon with characteristic size and period that matches the observed wave structures and could form the basis for physics based simple delay screen modelling.
- Fitting a simple 2nd order polynomial leaves residual structures with sizes of ~ 20 km and peak to peak (pp) variation of 0.005 TECU, corresponding to 2" pp position shifts at 100 MHz.
- The latter residual could be considered as effects of Kolmogorov turbulence induced by the medium scale TEC variations that dissipate by propagation into finer scales causing differential delay variation
- The occurrence of medium scale TIDs requires self-calibration at intervals of $10 - 10^2$ s and appropriate modelling.
- Although not discussed in detail the observed interferometer phase is also related to differential Faraday rotation and 0.1 TECU difference on longer

baselines could give ~ 1 rad differential rotation at 35 MHz as shown in subsection 4.1.2.

These results have an important impact on the design of a telescope, especially the choice for station beam size in relation to the self-calibration of a synthesis array and will be further discussed in more detail in subsequent sections. Our conclusions are:

- Refraction by a TID causes position shifts by the delay gradients, and also delay curvature causing image blur for an array larger than $1/12^{\text{th}}$ of a typical TID wavelength of ~ 90 km. Arrays larger than ~ 7 km need therefore for all stations further away than ~ 7 km from the centre of the array proper corrections for excess pathlength in each station beam as function of direction.
- To allow global modelling of a TID induced phase screen over the beam of each station with only 5 reference sources a limited beam size is required. A more detailed analysis will be given in subsection 4.8.2 showing that 5° might be sufficient.
- This fixed maximum beam size is in contrast with beam matching to the characteristic coherence size according to the Kolmogorov model that would require at longer wavelength a maximum beam width that scales almost proportional to frequency to make the small-scale phase deviations per telescope independent of frequency.

4.4 Multi-source self-calibration approach

This section summarizes the self-calibration approach that has been adopted for LOFAR [Noordam, 2006]. Conventional calibration uses strong reference sources in an almost empty field to estimate instrumental parameters or uses external means such as GPS for estimating ionosphere TEC to derive corrections for refraction and for Faraday rotation. It was realized early in the conceptual design phase [Bregman, 1998, 1999] that calibration would be a key issue, and LOFAR has been designed to be sensitive enough to rely completely on self-calibration using a number of sources inside and outside an observed field.

Current imaging packages use an iterative approach that starts with initial calibration parameters derived from calibration observations. These parameters allow crude imaging and identification of the strongest sources in the field, of which the strongest one is used for self-calibration that solves for varying complex gain factors per station. These gain factors are used to correct the visibilities and to subtract the identified sources in the field. In a next iteration step the gain factors are improved by reduced distortion of the strongest objects, an improved image is made and the next set of strongest sources is identified for subsequent subtraction. The subtrac-

tion process uses accurate correction for non-planarity, while 2-D Fourier imaging provides distorted object images.

In contrast with current imaging the LOFAR calibration pipeline already knows the sources that need to be subtracted, as they are available from a catalogue, the Global Sky Model [Nijboer, 2006]. The strongest 5 to 10 sources in the visibility data that are spatially filtered by the station beam (including the strong sources in station side lobes) are used for multi-source self-calibration. These so-called Category I (Cat I) sources are used to solve for phase and gain in at least five directions within the main beam of each telescope. A subsequent interpolation scheme then allows to use first-order estimates for the complex gain correction for all sources that are not strong enough to perform an adequate self-calibration.

Although the reference (point) sources can be subtracted accurately from the visibility data and leave no spurious responses in a synthesis image, the sources that use interpolated complex gain factors cannot be subtracted accurately and leave responses that ultimately could limit the sensitivity of a synthesis image. It has been shown by simulation [Tol, 2007] that a sky model containing 6 sources and noise can provide bias free estimates for the complex gain in 6 directions for each of 30 stations in the modelled synthesis array. Unfortunately this is not a proof that we get bias free estimates if in addition to Gaussian noise more weaker sources are present that cannot be solved for, but instead could disturb the solution of the stronger ones. Nevertheless the proposed method has been implemented [Tol, 2009] and named SPAM (Source Peeling and Atmospheric Modelling) [Intema, 2009] and shows clear improvement of two VLSS [Cohen, 2007] images where more than one source is strong enough to provide input for a phase screen model of the ionosphere.

There is a fundamental limitation in the number of sources for which a complex gain factor can be solved.

- Only less than $\frac{1}{2} (N_s - 1)$ complex source gains per beam can be solved for each of the N_s stations in a narrow band snapshot observation. This number is set by the mathematical limitation that no more independent parameters can be solved from a set of $\frac{1}{2} N_s(N_s - 1)$ independent complex visibilities while actual solving algorithms produce even less parameters depending on the SNR of the sources.
- Only longer observations with larger bandwidth provide more independent U,V-samples that allow in principle solving for more source gains, which is important when sub arrays are formed that have fewer stations.
 - For instance a 3 km interferometer of two parabolic tapered 40 m stations has a U,V-sample with an effective diameter ~ 31 m that contains only after 140 s tracking time independent sky information, while a 30 km baseline refreshes information after 14 s.
 - An interferometer sample with 32 m effective diameter and 3 km baseline, gives for every 1% relative frequency step new information.

Since for LOFAR the complex gain is generally known over a sufficient frequency range, it is proposed here for purposes of discussion to model the ionosphere not by a phase screen but by a curved thin delay slab over the synthesis array using delay estimates for at least five directions from every telescope in the array. This approach avoids not only 2π ambiguities encountered in construction of a curved phase screen but also allows estimation for each frequency of the proper phase for rays that intersect the slab with different inclination.

The proposed multi-source self-calibration has a number of requirements to be fulfilled and the first two are:

- A minimum effective station collecting area and receiver bandwidth to provide sufficient signal to noise ratio (SNR) for the 5 strongest sources in the station main beam. An $\text{SNR} > 3$ is required for as many baselines as needed to estimate essential parameters like the pointing offset of individual telescope beams and tilt and curvature of the ionosphere delay screen for each station beam for each ionosphere coherence time.
- The station main beam should be narrow enough to allow accurate subtraction of the next set of so-called Category II (Cat II) sources using the solved curvature of the ionosphere delay screen. Cat II sources are defined as having a $\text{SNR} > 3$ in a snapshot image and the weakest one is therefore a factor $\sim N_s^{-1}$ weaker than the weakest Cat I source.

The Cat I sources can now be defined more precisely as those sources that provide an $\text{SNR} > 3$ on a number of baselines from each telescope when observed with $\sim 20\%$ relative bandwidth within an ionosphere coherence time of order 10 seconds.

Apparently, the distinction between Cat I and Cat II is not based on intrinsic source properties but on the characteristics of the observing array. The 20% relative bandwidth is somewhat arbitrary, but provides sufficient coverage to separate ionosphere excess delay from sampling clock delay that is for LOFAR different for each telescope. It is precisely the limited bandwidth of 1.5 MHz of the 74 MHz receiver system and the very low aperture efficiency of the antenna system of $\sim 15\%$ on the 25 m dishes of the VLA that precluded the use of self-calibration on arbitrary sky fields. Instead the so-called field based calibration method has been developed for the VLA Low-frequency Sky Survey (VLSS) [Cohen, 2007]. The field based calibration uses the sensitivity of the full array to estimate a tip-tilt correction for every integration interval of 2 min, but does not correct for curvature in the phase screen over the array. The latter would require at least one Cat I source in the observed field and the assumption of receiver stability over the period since the last calibration on a reference field with a strong source that defines the instrumental station phases and which includes an arbitrary reference phase slope and phase curvature over the array. Unfortunately there are only few fields in the VLSS with at least one Cat I source, while for a survey a calibration procedure is required that is consistently used for all fields. The result is that the error side lobes of the Cat II sources,

which can no longer be accurately subtracted, will raise the noise floor in a snapshot image. Also the noise in longer synthesis image that is the average of a set of snapshot images will be increased and is investigated in section 5.3.

Already two requirements for the station performance have been mentioned and the third one is:

- The station side lobes should be sufficiently low such that no more than 3 sources over the sky outside the station main beam would require a self-calibration solution and absorb a fraction of the maximum number of solvable parameters per station beam. A sub-array with 17 stations could then still solve for the 5 strongest sources in the main beam and allow determination of an independent curved delay screen over each station for a narrow band snapshot image.

Quite luckily, the three strongest sources in the Northern hemisphere, Cas A, Cyg A and Tau A, are about a factor 10 stronger than should be expected from the source density on the sky as function of source intensity and about a factor 100 stronger than the average source in an average LOFAR station beam. This means in the first place that individual element antennas in the phased array stations can be calibrated using limited bandwidth and integration time [Wijnholds, 2011]. It also means that these three strongest sources are suppressed by the side lobes of a station with ~ 100 antennas (giving side lobes of $\sim 1\%$), but still strong enough to be solved for and subtracted accurately. However, all weaker sources fall below the limits defined for the Cat II sources as determined by the number of stations in the array. This is especially true when source attenuation due to time and bandwidth smearing of sources further away from the main beam is taken into account [Wijnholds, 2008]. This makes clear that only sources in near-in station side lobes could qualify as potential Cat II ones that need to be subtracted and need a proper complex gain estimate. For stations in the core of the array we find a delay screen that extends over the core and could in principle provide such phase corrections. For remote stations it is not possible to derive a proper phase correction for the Cat II sources in the near-in side lobes and it is important to reduce their apparent flux by applying an appropriate taper over the station array to reduce the near-in side lobes.

Although the source subtraction from the visibility data in the U,V,W-domain could in principle be done perfectly for the Cat I sources, the Cat II sources cannot be subtracted perfectly since the modelled curvature of the ionosphere phase screen and the predicted shape of the station beam are only first-order approximations. Especially for the lowest frequencies where LOFAR could operate, the station beam is larger than the ionosphere patch, which excludes accurate modelling with only 5 sources. Consequently, at the lowest frequencies, there is a subset of the Cat II sources in the main beam, for which the error side lobes are larger than the thermal noise floor of the snapshot image and these will effectively increase the snapshot noise.

Averaging of the snapshot images in a synthesis image lowers the thermal noise, the side lobe noise and also the error side lobe noise all in the same proportion since they are all three independent from snapshot to snapshot. The noise term that dominates the snapshot noise will also dominate the noise in a synthesis image. However, a weaker term that is correlated between snapshots averages away more slowly and could even dominate in the final synthesis image.

If the actual ionosphere coherence time is longer than the assumed 10 s then more sources could be solved for per snapshot and a more accurate delay screen could be constructed. However, when complex gain solutions are made for M source directions for each station in a narrow band snapshot, there are at most $\frac{1}{2} (N_s - 1)$ complex visibilities per station that are strong enough. A least squares fit for M independent source fluxes and positions leaves at most $\sim (\frac{1}{2} (N_s - 1) - M)$ independent complex visibility noise contributions that determine the final noise in the solutions for direction dependent complex gain factors. This means that using the fitted solution introduces a station based complex gain error for the Cat II sources with a SNR that depends on the SNR of the fitted Cat I sources, on the number of remaining independent visibility contributions and on the disturbing contribution of the Cat II sources that are not solved for. This issue will be discussed further in section 4.7.

Although systematic errors could be reduced by including more sources and by solving for more directions, the statistical error is increased by leaving too few independent baselines. This indicates the importance of an array with sufficient stations that are large and have a beam narrow enough to limit M to a value that is still sufficient to describe the ionosphere delay screen sufficiently accurate [Wijnholds, 2011].

This calibratability requirement is different from the requirement to provide sufficient U,V-coverage, which drives to many small stations.

We have shown that two related issues determine the feasibility of multi-source self-calibration, as the approach that will provide sky noise limited sensitivity performance. These issues are, (i) sufficient sources in a station beam of sufficient strength to solve for delay screen parameters, and (ii) interpolation with these parameters to provide phases that are accurate enough to subtract the strongest sources in the field such that their residual artefacts do not spoil the thermal noise floor of an image. These questions will be addressed further in sections 4.6 and 4.8 respectively.

4.5 Angular density of sources as function of their flux and size

Calibration of a synthesis array according to the procedure defined in subsection 4.4 needs modelling of a curved delay screen over the beam of remote LOFAR stations, which requires detection of at least 5 calibration sources of sufficient strength in the station main beam. In section 4.6 we will show that a source strength > 0.1 Jy at 140 MHz for 10 s sampling is needed for ionosphere phase correction, but beam shape calibration could easily integrate for 10^3 s reaching many more sources > 0.01 Jy that should not be resolved by the longest baselines. European stations at distances up to 600 km from the core of LOFAR require sources or source components that are only partially resolved by interferometer resolutions of $0.7''$ to $3''$ at 140 MHz and 35 MHz respectively. However, calibration of baselines towards stations out to 80 km from the core could even use sources up to $5''$.

In addition to integrated source count formulae for the relatively strong calibration sources, to limit the noise from artefacts, as discussed in chapter 5, one needs estimates for the total number of sources at the $1\sigma_{\text{rms}} < 0.01$ mJy noise level that could ultimately define the effective noise floor of deep LOFAR images by side lobe confusion.

This section derives approximate cumulative source count formulae from published differential source count data and discusses the range of applicability.

4.5.1 Introducing cumulative and differential source counts

The cumulative (also-called integrated) source count $N(>S)$ gives the number density of objects per steradian (sr) that are stronger than threshold flux S . From a sky image observed with a synthesis radio telescope we can extract sources and find their fluxes, but limited resolution of the instrument could merge a number of unresolved sources to a single one. On the other hand an extended object that is resolved could have a peak flux well below the detection threshold while integration over a number of resolution elements provides a proper detection. A further complication is that there are different populations of objects each with characteristic morphology and different radiation mechanisms for their apparently constituting components that each have a characteristic maximum luminosity (W Hz^{-1}). Finally each population decreases not only in observed intensity with increasing distance, but at cosmological scales, objects further away belong to a different epoch where physical conditions could be different for different populations. The population issue is illustrated in figure 4.8 where a model for the source count is presented using two populations.

The large range in flux and number density is conveniently represented in a log-log graph where power law relations show up as straight lines. The cumulative source

count for a single population with homogeneous distribution in Euclidian space would be given by $N(>S) = N_0 (S/S_0)^{-1.5}$ and this function is often used to normalize observed data. For astronomical analysis it is more convenient to work with the derivative of the cumulative source count and normalize by $S^{-2.5}$ to get the Euclidean normalized differential source count.

Astronomical literature has a focus on deriving intrinsic astrophysical properties but our focus is on deriving first-order estimates of relevant observational parameters in a way that can be used for non-astronomical purposes. Such results can be easily verified once LOFAR is fully operational, but it will take a long time before relevant data will be published as is shown by comparing the dates of our reference publications with the dates when the used instruments became operational. Combining low sensitivity and low resolution observations of large fields in the frequency regime 20 – 200 MHz with higher sensitivity or higher resolution observations in small fields at higher frequencies leads to useful results when some simplifying assumptions indeed hold.

Important assumptions are that (i) the observed fields are representative for the sky as a whole and (ii) the same source populations are compared. We use observed source counts, observed spectral index information, and observed size-flux density relations, but avoid cosmic evolution issues that are not well established.

Figure 4.8 shows that the normalized differential source count can be described to first-order by three straight lines in a log-log graph, a lower and an upper plateau and a connecting slope.

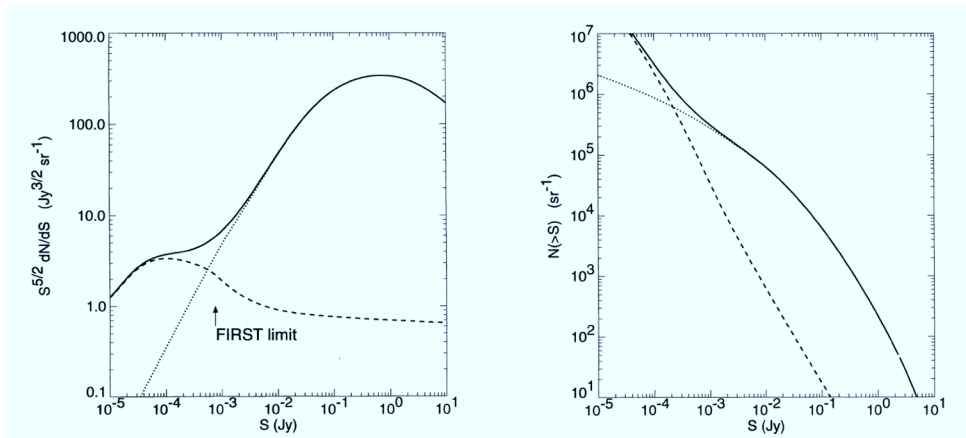


Figure 4.8. Replication of fig. 2 [Becker, 1995]. (left) Differential source number density, dN/dS , vs. flux density S at 1.4 GHz. Short and long dashes represent the modelled contributions from AGNs (Active Galactic Nuclei) and star forming galaxies, respectively. (right) Relative contributions to the cumulative source number density $N(>S)$.

A more refined description needs the connecting slope to be broken up in at least two but preferable four parts. The most sensitive observations determine the lower plateau level and the flux from where an upward slope starts. From the large field observation the upper plateau level is determined and a downward slope, which intersects the upward slope around ~ 0.04 Jy in the graphs for 1.4 GHz. A comparable break in slope is expected around ~ 0.3 Jy for 140 MHz and at ~ 1.1 Jy for 35 MHz if a spectral index of 0.9 is valid. For beam shape calibration we can integrate for 10^3 s and then need for $\text{SNR} > 3$ sources > 1 Jy at 35 MHz and sources > 0.01 Jy at 150 MHz. These sensitivities are a factor ten lower than for phase calibration at 10 s intervals, and as a consequence, the number of sources per station beam that are strong enough for self-calibration varies as function of sensitivity for 35 MHz, 70 MHz and 150 MHz.

4.5.2 Analysis of source counts at 38, 151 , 325 and 1400 MHz

We start looking into available survey data and spectral indices at different flux levels relevant for self-calibration.

The 8C survey ($\delta > 60^\circ$) [Rees, 1990] observed with the Cambridge Low-Frequency Synthesis Telescope (CLFST) at 38 MHz has a limiting source sensitivity of 1 Jy ($5\sigma_{\text{rms}}$). The (CLFST) is an almost East-West array of 4.6 km length that has a resolution of $4.5' \times 4.5' \text{ cosec}(\delta)$ at that frequency.

Comparing the flux of a small 8C sample of 57 sources in an area of 29 square degrees around the Ecliptic Pole with the flux at 151 MHz, showed a median spectral index $\alpha_m = 0.8$ for sources stronger than 1.3 Jy [Lacy, 1992], which is equal to the spectral index derived from comparison with 4850 MHz.

The revised 3C catalogue [Bennett, 1962] covers $\delta > -5^\circ$ and contains 330 sources stronger than 10 Jy at 178 MHz, which allows estimating the contribution to observed visibilities by objects in the side lobes of the station beam. The integrated source density has in a $\log N(>S)\text{-log } S$ graph a slope index -1.9 which is still steeper than -1.5 for a homogeneous, static, Euclidean Universe.

The 7C survey observed with the CLFST at 151 MHz, is more sensitive and has a higher resolution of $70'' \times 70'' \text{ cosec}(\delta)$ but consists of various parts. One part has two regions covering ~ 0.144 sr around RA $8^{\text{h}} 28^{\text{m}}$ DEC 43° including the Lynx area and contains 4723 sources > 0.08 Jy ($5\sigma_{\text{rms}}$) [McGilchrist, 1990].

The normalized differential source density has a plateau at $3600 \text{ (Jy}^{1.5} \text{ sr}^{-1})$ for fluxes between 1 - 10 Jy and decreases with slope index 0.58 for fluxes below 1 Jy. About 90 % of the sources are not resolved and comparison with 408 MHz observations shows a spectral index distribution with median $\alpha_m = 0.90$ (for $S = S_0 (\nu/\nu_0)^{-\alpha}$), which is roughly equal to the distribution derived from comparison with 1.4 GHz data.

GMRT observation of the Lynx field at 150 MHz [Ishwara-Chandra, 2010] provided 765 sources in a field of 15 deg^2 . The derived normalized differential source count has a slope 0.99 for the whole range from the limiting magnitude at $\sim 4 \text{ mJy}$ ($6\sigma_{\text{rms}}$) up to $\sim 1 \text{ Jy}$. About two-thirds of the sources are unresolved by the resolution of $19'' \times 14''$. Spectral indices for the large field were derived using various catalogues such as the NVSS at 1.4 GHz that has a resolution of $45''$ and a limiting sensitivity of $\sim 2.5 \text{ mJy}$ ($5\sigma_{\text{rms}}$) for sources at $\delta > -40^\circ$ [Condon, 1998]. Combining the GMRT data with these and other deeper observations show a spectral index distribution based on 639 sources (83%) with a median value of $\alpha_m = 0.78$. A closer look shows that this median actually changes from ~ 1.0 at 200 mJy to ~ 0.6 at 10 mJy .

The Westerbork Northern Sky Survey (WENSS) [Rengelink, 1997] covers $\delta > 30^\circ$ at 325 MHz with a resolution of $54'' \times 54'' \text{ cosec}(\delta)$ by the 2.8 km EW array. The relevant part of the normalized differential source count can be described by a plateau at $2000 (\text{Jy}^{1.5} \text{ sr}^{-1})$ for $6 > S_{325} > 0.7 \text{ Jy}$ and a decreasing slope with index of 0.66 for $0.7 > S_{325} > 0.03 \text{ Jy}$ close to the survey limit at 18 mJy ($5\sigma_{\text{rms}}$) where the slope steepens.

A deeper survey of the Lynx field at 327 MHz [Oort, 1988] reached a limit of 4.5 mJy ($5\sigma_{\text{rms}}$) where image noise starts to become dominated by the confusion noise of $\sim 0.6 \text{ mJy}$ (σ_{rms}). A decreasing slope with index 0.64 was found for the same interval as found by WENSS and steeping to slope index 0.88 below 30 mJy . An important result was the decrease in median spectral index relative to 1.4 GHz from ~ 0.7 at $\sim 30 \text{ mJy}$ to ~ 0.5 at $\sim 6 \text{ mJy}$. These results agree very well with the 150 MHz observations of the same Lynx field presented above [Ishwara-Chandra, 2010].

The FIRST survey at 1.4 GHz has a resolution of $5.4''$ and reaches $\sim 1 \text{ mJy}$ ($7\sigma_{\text{rms}}$) [Becker, 1995]. A catalogue from the initial 1550 deg^2 contains 138,665 sources [White, 1997]. The points in the graph of the normalized differential source count shown in figure 4.9 can be described accurately by three straight lines with slope index 0.36 for $0.3 > S_{1.4} > 0.1 \text{ Jy}$, with index 0.69 for $0.1 > S_{1.4} > 0.02 \text{ Jy}$ and with index 0.82 down to 2 mJy .

The WSRT 1.4 GHz amalgamated source count [Katgert, 1988] has a highest resolution of $19''$ and shows a plateau of $340 (\text{Jy}^{1.5} \text{ sr}^{-1})$ in the normalized differential source count for $1.7 > S_{1.4} > 0.3 \text{ Jy}$ and a decrease with slope index 0.56 to 0.01 Jy . A slope with index 0.95 is valid for $20 > S_{1.4} > 1 \text{ mJy}$ and a slope of 0.3 down to the limit at 0.1 mJy ($5\sigma_{\text{rms}}$).

These source counts are heavily weighted in the data given by [Windhorst, 1990] that focussed on the turn up below 1 mJy . The resulting smooth fit that averages over a number of noisy surveys describes the turnover at $\sim 0.3 \text{ Jy}$ and the turn up at $\sim 1 \text{ mJy}$ as indicated by the shaded curve in figure 4.9 but masks the break at $\sim 20 \text{ mJy}$.

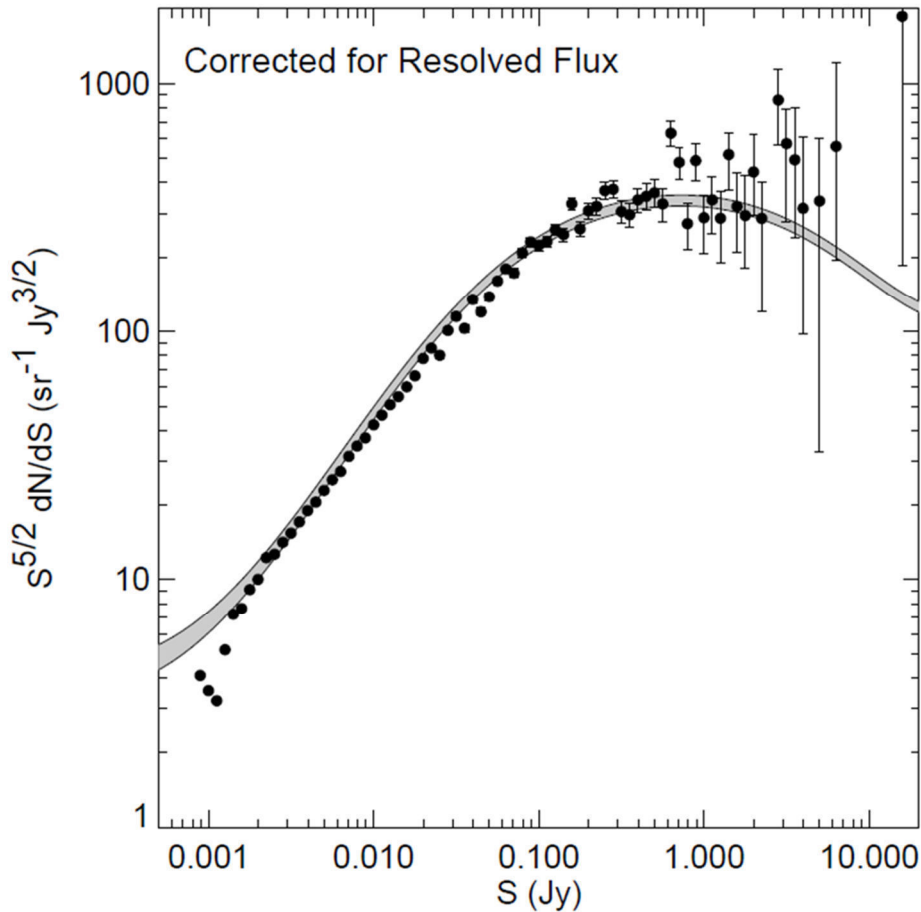


Figure 4.9. Replication of figure 11 [White, 1997] representing normalized differential source counts from the FIRST survey at 1.4 GHz that are corrected for resolved flux. A break in slope at 20 & 100 mJy is visible but smoothed away in the shaded curve provided by [Windhorst, 1990].

4.5.3 Source sizes at 20 cm and 90 cm and suitability as LOFAR calibrators

An excellent overview on the cosmic evolution of weak radio galaxies by [Windhorst, 1990] summarized available source count data and produced figures of the median source size as function of 21 cm flux. An important relation for our self-calibration requirement is that the observed median source size Θ_{sz} can be described as power law function of flux $S_{1.4}$ in mJy at 1.4 GHz by

$$\Theta_{sz}(S_{1.4}) = 2 S_{1.4}^{0.3} \quad [\text{arcsec}] \quad \text{for } 10^{-1} < S_{1.4} < 10^3 \text{ mJy} \quad (4.22)$$

In fact there is a bi-modal distribution showing that 22% of the sources in the range $3 < S_{1.4} < 100$ mJy is smaller than 1" and 37% is smaller than 3".

The cumulative distribution of all sources in the FIRST catalogue presented in figure 5 of [White, 1997] shows that 80% of all objects is smaller than 5.4" consistent with an integrated version of (4.22). This means that there is no lack of sources suitable as calibrator on baselines up to 80 km for frequencies below 200 MHz.

Unfortunately, a statement that 22% of the sources with $S_{1.4} < 100$ mJy is smaller than 1" is not sufficient to establish whether sources with $S_{150} > 100$ mJy are smaller than 0.5" which is required to provide sufficient signal on baselines of ~600 km between the LOFAR core and European stations.

To get an impression of availability of suitable calibrators two fields 1.9° apart were observed at 324 MHz [Lenc, 2008] each covering 3 deg^2 with effective resolutions between 0.2"- 0.4". The two fields together contain 50 sources from the WENSS catalogue [Rengelink, 1997] stronger than 80 mJy that could show unresolved components given the sensitivity of the high resolution observations. Only 14 sources are found smaller than 0.5" and 3 sources extend to 4", all showing up with one to three resolved components. The sources $< 0.5''$ are distributed over three bins defined by 80-160, 160-320 and >320 mJy for their integrated flux which results in a cumulative source count that follows the distribution of the WENSS sources at a level that is a factor four lower. The integrated component flux originates from WENSS sources that are a factor 1.3 stronger on average, but one source is nine times stronger.

These results are independent of the spectral index of the WENSS sources with detected compact components. The spectral index distribution is bi-modal with 4 sources having a mean spectral index of 0.18 and 10 sources with a mean index of 0.89 giving an average of 0.68 over the bi-modal distribution. This value is comparable to 0.8 as found for the spectral index of objects stronger than 1.3 Jy at 38 MHz [Lacy, 1992]. Their sample contained only 4% sources with a low spectral index.

Also [McGilchrist, 1990] found a higher average spectral index of 0.9 for their 151 MHz objects stronger than 0.08 Jy, but their sample contained only 2% low index sources.

The 57 sources in the 38 MHz sample by [Lacy, 1992] were imaged at ~5 GHz and 25% turned out to be smaller than 5" and having even smaller components, consistent with the results from [Lenc, 2008].

The results of this subsection can now be summarized as follows

- The spatial properties of objects stronger than 0.08 Jy at 324 MHz are also valid at frequencies as low as 38 MHz, hence appropriate source counts can be derived from source counts at higher frequencies by adopting an averaged spectral index $\alpha_{xx}^{1.4} = 0.8$ between 1.4 GHz and frequencies xx down to 30 MHz.
- In particular we conclude that these spectral indices are also appropriate to transform the size-flux correlation at 1.4 GHz to lower frequencies to define for $S_{140} < 1$ Jy a 37% subclass of objects smaller than 3", and a 22% subclass with objects smaller than 1" based on the bi-modal size distribution given by [Windhorst, 1990] instead of (4.22).
- From the data presented by [Lenc, 2008] we derived that objects stronger than 80 mJy at 324 MHz and smaller than 0.5" have up to three even smaller components and constitute a 25% subclass of the cumulative source count.

4.5.4 Source properties below 1 mJy

The previous paragraphs addressed the range of source fluxes and source sizes relevant for self-calibration of LOFAR observations. In the next paragraphs we address the number of sources that could appear at the lowest sensitivity levels in long and repeated synthesis observations and could finally set a confusion limit.

Earlier work using optical identifications concluded [Oort, 1988] that the upturn in the Euclidian normalized differential source density below 0.8 mJy at 1.4 GHz is to be attributed to a previously unsuspected blue galaxy population. Other well explored fields are the Hubble Deep Field (HDF) and Hubble Flanking Fields (HFF) for which deep ($1 \sigma_{\text{rms}} \sim 8 \mu\text{Jy}$) WSRT observations have been made at 21 cm [Garrett, 2000]. The introduction of the latter paper summarizes "that ~60% of the faint sub-mJy sources are star forming blue galaxies with steep HII-like emission spectra at moderate distance ($z \sim 0.2 - 1$). The remaining ~20% of the faint radio population are identified with relatively low-luminosity Active Galactic Nuclei (AGN) and ~20% have no visible optical counterpart".

Recent deep 1.4 GHz observations covering a region of the SWIRE Spitzer Legacy survey went down from 0.9 to 0.015 mJy ($\sim 5\sigma_{\text{rms}}$) [Owen, 2008] and show even a slight increase in the normalized differential source count from 4.5 - 6.5 [$\text{sr}^{-1} \text{Jy}^{1.5}$] respectively. Interestingly, all sources < 1 mJy have median angular sizes $\sim 1.2''$ as observed with a VLA resolution of $1.6''$. The same field observed at 324.5 MHz [Owen, 2009] reached $5\sigma_{\text{rms}} \sim 0.4$ mJy per beam of $6''$ and shows a flat part in the normalized differential source count below ~ 2 mJy.

This value is consistent with a value of 0.8 mJy for the start of the flat part in the 1.4

GHz source distribution using the peak in the spectral index distribution at 0.7 as derived for these sources.

The decrease of the spectral index with lower flux observed at 150 MHz [Ishwara-Chandra, 2010] for $S_{150} < 200$ mJy is also observed at 324.5 MHz for sources with $S_{1.4} < 10$ mJy and sizes $< 3''$ having $\alpha_m \sim 0.3-0.5$. However, for sources with $S_{1.4} < 1$ mJy and sizes $< 3''$ there is a trend to higher spectral indices for lower fluxes. It is finally concluded by [Owen, 2009] “that the changing spectral index of the sources with $S_{1.4} < 1$ mJy is not well understood but probably involves the Active Galactic Nuclei population”, which indicates a difference of opinion with [Garrett, 2000] cited above.

Detailed astronomical consideration is outside the scope of this discussion, but the summarized material allows at least to assume that below 0.1 mJy (at 1.4 GHz) the sky is dominated by galaxies having angular sizes $< 1.6''$ with spectral index < 0.7 that will define a side lobe confusion limit for LOFAR, to be discussed in chapter 5.

4.5.5 Deriving 1.4 GHz cumulative source count and frequency scaling formulae

The characteristics of the various normalized differential source counts are summarized in table 4.1.

Table 4.1. Characteristics of various normalized differential source counts.

(dN/dS) $S^{2.5}$	units	150 MHz	325 MHz	1.4 GHz
High plateau level	$\text{Jy}^{1.5} \text{sr}^{-1}$	3600 ^3	2000 ^4	$340 \text{ }^{1.5}$
High plateau range	Jy	$1 - 10 \text{ }^3$	$0.7 - 6 \text{ }^4$	$0.3 - 1.7 \text{ }^{1.5}$
Down slope index		0.58 ^3	$0.66 \text{ }^4, 0.64 \text{ }^2$	$0.36 \text{ }^5, 0.56 \text{ }^1$
Intersect	Jy	0.19 ^3	$0.07 \text{ }^7, 0.04 \text{ }^2$	$0.06 \text{ }^5, 0.02 \text{ }^1$
Up slope index		0.95 ^8	$0.88 \text{ }^7, 0.95 \text{ }^2$	$0.82 \text{ }^5, 0.95 \text{ }^1$
Low plateau turn up at	mJy	-	2.2 ^7	$0.8 \text{ }^{1.6}$
Low plateau level	$\text{Jy}^{1.5} \text{sr}^{-1}$	-	21 ^7	5 ^6

¹ [Katgert, 1988], ² [Oort, 1988], ³ [McGilchrist, 1990], ⁴ [Rengelink, 1997], ⁵ [White, 1997], ⁶ [Owen, 2008], ⁷ [Owen, 2009], ⁸ [Ishwara-Chandra, 2010]

Interestingly, we find different intersects of up and down slope by using different datasets at 325 MHz as well as at 1.4 GHz. A closer look at the 150 MHz data [Ishwara-Chandra, 2010] reveals an anomaly around 0.1 Jy that could be attributed to

the Lynx field and using a spectral index 0.9 we find that anomaly back in the Lynx fields at 325 MHz [Oort, 1988] and at 1.4 GHz [Katgert, 1988].

Moreover the average of a number of datasets still shows a knee at 20 mJy that is also visible in the FIRST data [White, 1997] of figure 4.9. The latter data follow a perfect straight line from 2 – 20 mJy but the region 20 – 80 mJy shows a variance between the samples that is much larger than the statistical variance in the samples. Surprisingly, the corresponding 75 – 300 mJy region over a three times smaller area at 327 MHz [Rengelink, 1997] does not show this additional variance. A straight line from 30 – 700 mJy gives a better fit to the data than the curve shown in their figure 11 that also has to describe the turnover at 2 Jy with only a few parameters. In the same vein the smooth curve by [Windhorst, 1990] shown shaded in figure 4.9 masks the knees visible at 20 mJy and at 100 mJy.

Instead of a normalized differential source density we need integration to get the cumulative source density $N(>S)$ for all sources stronger than S as function of S . The data at 1.4 GHz are the most accurate and proper spectral indices between 1.4 GHz and 150 MHz have been discussed in a previous subsection for $S_{1.4} > 2$ mJy. For the lower flux levels we use values derived by [Owen, 2009] from stacking 6" resolution data at 324 MHz for sources $< 3''$ at 1.4 GHz. It has been shown [Lacy, 1992] that the spectral index between 150 MHz and 38 MHz is ~ 0.8 for $S_{38} > 1.3$ Jy and we assume a slightly lower value 0.7 for lower flux levels.

To simplify the integration we define seven flux ranges and describe the log-log graphs of the published differential source density with seven properly matched straight lines. It might be argued that such a description could be over-interpretation of the data and that a smooth fit is more natural. Although the choice of knees is indeed somewhat arbitrary and biased towards the important calibration regimes at 150 MHz, it exaggerates the effects of passing a knee but fits perfectly within the accuracy range of the data.

We combine the 1.4 GHz data of [Windhorst, 1990] for $S_{1.4} > 0.3$ Jy (slope index - 0.4 for $S_{1.4} > 1.7$ Jy) with the slope data from [White, 1997] for $S_{1.4} > 2$ mJy. Although the deepest data from the Swire field suggest an upward slope for $S_{1.4} < 0.2$ mJy [Owen, 2008], other datasets in their figure 11 still show a decrease and we will use a conservative value of $5.5 \text{ (Jy}^{1.5} \text{ sr}^{-1})$ below 0.6 mJy. For $0.6 < S_{1.4} < 2$ mJy we use the shaded graph in figure 4.9 based on data by [Windhorst, 1990] giving a slope 0.56.

In table 4.2 we have selected appropriate intervals to fit straight lines in the published graphs to derive dN_r/dS for the various ranges and defined range boundaries marked with a $>$ sign. The column with $N_r(>S)$ gives the integration of dN_r/dS over S from S to infinity.

To get the proper numerical result for each $N(>S_r)$ we need to subtract the numerical value $N_r(S_{r \text{ upper}})$ valid for the upper boundary of the interval and add the lower boundary value $N_r(S_{r \text{ lower}})$ of the next higher interval according to the integration formula where

\int_S^S indicates integration between boundaries S and s .

$$\begin{aligned} N(>S) &= \sum (dN/dS) dS \\ &= \int_S^S dN_1 + \int_S^S dN_2 = N_1(>S) - N_1(>s) + N_2(>s) - N_2(>\sigma) \end{aligned} \quad (4.23)$$

We start therefore at the highest interval that has zero upper boundary value ($N_2(\sigma) = 0$ for large σ) and needs no correction and then calculate successively the correct values $N(>S_r)$ for all the lower boundaries where the slope changes. The intermediate results are not shown in the table only the final numerical values of the integrated source count at each boundary in column $N(>S_r)$.

Table 4.2. 1.4 GHz differential and cumulative source counts and scaling to 140, 70 & 35 MHz

S_r	dN_r/dS	$N_r(>S)$	$N(>S_r)$ sr^{-1}	$N(>S)$ sr^{-1}	α_{men}	S_{140}	S_{70} $_{7,8}$	S_{35} $_{7,8}$
>0.02) ¹	$5.5 S^{-2.5}$	$3.67 S^{-1.5}$	$4.12 \cdot 10^7$	$23.1 S^{-1.33}$	0.6) ⁴	0.08	0.13 0	0.21 1
0.06					0.5) ⁴	0.19	0.31	0.50
0.2 mJy					0.4) ⁴	0.50	0.81	1.32
>0.6) ²	$340 S^{-1.94}$	$362 S^{-0.94}$	$4.43 \cdot 10^5$	$2285 S^{-0.71}$	0.5) ⁴	1.90	3.09	5.02
>2.0) ³	$1881 S^{-1.68}$	$2766 S^{-0.68}$	$1.88 \cdot 10^5$	$1303 S^{-0.80}$	0.6) ⁴⁵	8.00	13.0	21.1
6.0					0.7) ⁵	30.1	48.9	79.4
>20) ³	$1117 S^{-1.81}$	$1379 S^{-0.81}$	$2.98 \cdot 10^4$	$594 S^{-1.00}$	0.8) ⁵	126	205	333
>100) ³	$528 S^{-2.14}$	$463 S^{-1.14}$	$5.92 \cdot 10^3$	$271 S^{-1.34}$	0.8) ⁶	631 mJy	1.10 Jy	1.91 Jy
>0.3 Jy) ²	$340 S^{-2.5}$	$227 S^{-1.5}$	$1.36 \cdot 10^3$	$192 S^{-1.63}$	0.8	1.89	3.29	5.72
>1.7 Jy	$423 S^{-2.90}$	$223 S^{-1.90}$	81	$223 S^{-1.90}$	0.8	10.7	18.7	32.5
>20 Jy			0.75		0.8	126	220	382

¹) [Owen, 2008], ² [Windhorst, 1990], ³ [White, 1997]

⁴) $\alpha^{1.4}_{324}$ [Owen, 2009], ⁵) $\alpha^{1.4}_{150}$ [Ischwara-Chandra],

⁶) 0.8 gives correct integrated source count at 150 MHz although $\alpha^{408}_{151} = 0.9$ [McGilchrist, 1990]

⁷) $\alpha^{151}_{38} = 0.8$ for $S_{38} > 1.3$ Jy [Lacy, 1992], ⁸) assume $\alpha^{140}_{xx} = 0.7$ for $S_{xx} < 0.1$ Jy.

Finally we calculate the slope in log-log coordinates between the numerical values of the cumulative source count for each interval and the appropriate factor of the power law $N(>S)$. For use at lower frequencies we need the mean spectral index as given by the references.

Interestingly, the slope index of the final integrated source count $N(>S)$ is not just the slope index of the differential source count plus one. We have now a set of formulae with less sharp knees having an integration error less than the statistical error in the data points at the knees. More important is that we have a smooth transition in the slope index of the integrated source count for the relevant flux regimes for self-calibration sources, for the relevant flux ranges (transformed to 1.4 GHz) of self-calibration sources for LBA and HBA array.

4.5.6 Conclusions

The conclusions of subsection 4.5 are

- An integrated source count covering 0.02 mJy to 20 Jy at 1.4 GHz has been constructed from published differential source counts while maintaining the statistical accuracy of the contributing segments.
- A spectral index of ~ 0.8 is appropriate to find integrated source counts for frequencies down to ~ 30 MHz for $S_{1.4} > 20$ mJy.
- For flux levels down to 0.2 mJy the spectral index flattens to 0.4 and increases to 0.6 again at 0.02 mJy.
- Analysis of published high resolution data at 324 MHz shows that sources with $S_{324} > 80$ mJy and size $< 0.5''$ constitute a 25% subclass that contains one to three even narrower components with a spectral index $\alpha_{324}^{1.4} \sim 0.7$.

4.6 Number of expected calibration sources per station beam

The number of sources per station beam that provide a signal to noise ratio (SNR) > 3 on a sufficient number of baselines with that station is a critical parameter that determines whether a delay screen can be determined for the station beam. In this section we consider the situation confronted by LOFAR.

Subsection 4.6.1 derives the sensitivity for a number of LOFAR stations at representative frequencies.

Subsection 4.6.2 derives the number of sources per station beam using the source density and spectral index derived in section 4.5.

Subsection 4.6.3 discusses how the spatial sampling for a delay screen can be improved by using adjacent beams that partially overlap the disturbing part of the ionosphere.

Subsection 4.6.4 summarizes the conclusions.

4.6.1 Sensitivity of LOFAR interferometers

The sensitivity of an antenna can be expressed as Source Equivalent Flux Density (S_{SEFD}) given by

$$S_{\text{SEFD}} = 2 k_B T_s A_e^{-1} \quad [\text{W m}^{-2} \text{ Hz}^{-1}] \quad (4.24)$$

with Boltzmann's constant k_B ($1.38 \cdot 10^{-23}$ [J K⁻¹]), system temperature T_s and effective aperture A_e . Taylor, [Taylor, 1999] gives the sensitivity for un-polarized flux ΔS of a single polarization interferometer formed by two equal antennas, where the source contribution to the system temperature can be neglected as

$$\Delta S = \eta_s^{-1} S_{\text{SEFD}} (2 B_e \tau)^{-1/2} \quad [\text{W m}^{-2} \text{ Hz}^{-1}] \quad (4.25)$$

With correlation efficiency $\eta_s \sim 1$ (for 12-bit correlation by LOFAR) effective bandwidth B_e and integration time τ .

For frequencies below 400 MHz the sky brightness is dominated by the galactic radiation which depends strongly on the wavelength. In practice we have to take into account that the actual system temperature T_s includes contributions from ground radiation T_g from receiver T_r and from the sky

$$\begin{aligned} T_{\text{sky}} &= 0.17 T_{150} \lambda^{2.55} && \text{with } \lambda \text{ in meters} \\ T_{150} &= 350 \pm 120 \text{ [K]} && \text{for galactic latitude } 10^\circ - 90^\circ \end{aligned} \quad (4.26)$$

A more detailed sky reference temperature T_{150} at 150 MHz is given in Figure 4.10 with an all sky overview of the actual sky brightness at 150 MHz that could be used to establish the sensitivity at a specific location.

For a phased array station with antenna receptors that have a beam solid angle Ω and where the N_r antennas operate in sparse mode, where λ is smaller than twice the separation between the antennas, we find an maximum antenna aperture A_m that excludes mutual coupling effects [Kraus, 1988] for observation in zenith direction equal to

$$A_m = N_r \lambda^2 / \Omega \quad \text{Typically } \Omega \sim 3 \text{ for LOFAR} \quad (4.27)$$

At larger zenith angles the effective area A_e shrinks proportionally to the directivity pattern of the element and varies from $\sim A_m \cos \theta$ to $\sim A_m \cos^2 \theta$ depending on azimuth.

For longer wavelengths and uniform element distribution we need, instead of A_m , the physical area A_p of the station and the effective area goes by $\sim A_p \cos \theta$ independent of azimuth. Actually, we need all individual element patterns from electromagnetic analysis as well as the full mutual impedance matrix of the array, with antenna impedances on the diagonal, to evaluate the matching of the antennas to the low noise amplifiers. This impedance matching determines the effective noise temperature of the receivers in the array and becomes a function of zenith angle. Also the contribution of the sky brightness temperature becomes a function of zenith angle since it needs integration over the full array pattern including side lobes and the grating lobes [Ivashina, 2008]. In the LBA where the receptors have an increasing density towards the centre we can as first-order estimate use the physical area as limited by half the distance towards nearest neighbours [Nijboer, 2009].

The effective width of the station beam is increased by tapering. We have so-called spatial taper by weighting the effective density of each receptor to decrease the level of the side lobes close to the main lobe.

1970AmJPA...16....1L

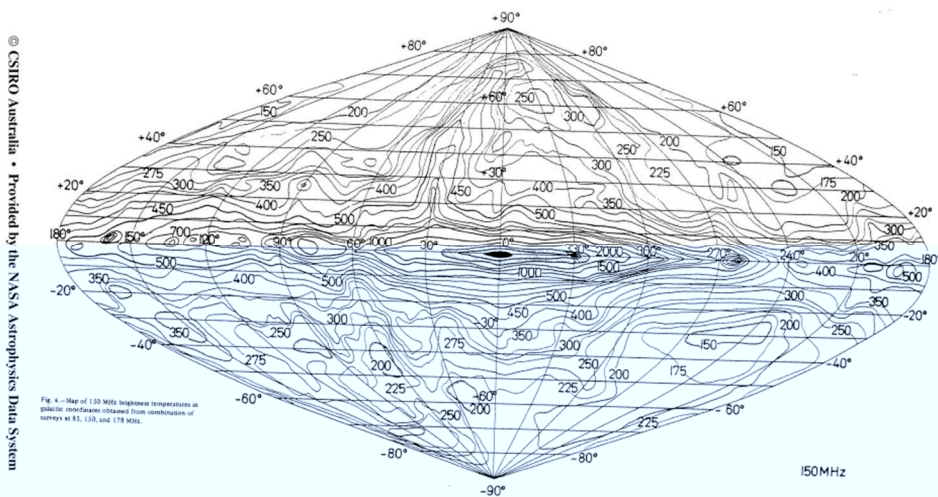


Figure 4.10. Sky brightness distribution in galactic coordinates at 150 MHz from surveys at 85, 150 and 178 MHz.

Electronic tapering reduces the relative weight of the receptor signals in the beam forming process and has the same beam effects but unfortunately reduces the effective area given by

$$A_{te} = \eta_{te} A_e \quad (4.28)$$

If a parabolic taper is used reaching zero at the edge of the aperture we get an electronic tapering efficiency $\eta_{te} = 0.75$ and an increased beam width $1.28 \lambda / D$ (FWHM) for a circular aperture instead of $1.01 \lambda / D$ for uniform illumination (for references see [Bregman, 2004a]). In survey applications the loss in sensitivity is almost fully compensated by an increase in survey area by the larger beam and leads only to a decrease of 5% in survey sensitivity. Such a marginal loss is to be preferred for deep surveys since an even larger loss could be possible by increased noise in synthesis images due to sources in the near side lobes that are far less reduced if no station taper would be applied.

A reasonable relative bandwidth is $\sim 20\%$ which allows simple linear approximations for parameters that change as function of frequency. Actual values are slightly different since the product of bandwidth per beam and number of beams is determined by the processing bandwidth of the correlation platform and by the requirement for an integer number of beams.

Table 4.3. Values of some LOFAR station properties and interferometer sensitivities

Frequency	MHz		35		70		140	
T_{sky}	K		14370		2454		419	
S_{EFD} per dipole*	MJy		2.75		1.88		1.29	
Effective bandwidth B_e	MHz		6		14		42	
Number of beams			7		3		1	
Station type		LBA	LBA _E	LBA _S	LBA _E	HBA _C	HBA _R	HBA _E
Dual pol antennas		46	96	46	96	24*16	48*16	96*16
sensitivity σ **	Jy	7.28	3.49	3.26	1.56	0.155	0.0773	0.0386
Equivalent diameter	m	81 ⁺	68 ⁺	32 ⁺	68 ⁺	28 ⁺⁺	40 ⁺⁺	56 ⁺⁺
$\theta_{1/2} = 1.28 \lambda/D$ FWHM	°	7.76	9.25	9.83	4.63	5.71	4.00	2.86
$A_m / A_{physical}$		0.22	0.63	0.35	0.16	~ 1	~ 1	~ 1

* uses $T_s = 1.7 T_{sky}$

** 1 σ_{rms} value per interferometer after 10 s including taper efficiency $\eta_{te} = 0.75$ assuming equal stations per interferometer

⁺ maximum antenna separation in LBA

⁺⁺ equivalent circular area filled with $N_t / 16$ tiles each providing $(5.15)^2 \text{ m}^2$

Table 4.3 summarizes some key parameters of the phased array stations used in LOFAR where subscript _E stands for the European, subscript _C for the core and subscript _R for the Dutch remote stations respectively.

Subscript _S is used for the small version of the low band array (LBA) that uses only the central part of the array. We use representative frequencies of 35, 70 and 140 MHz where the station arrays are indeed sparse. Unfortunately the European stations lose some sensitivity at 35 MHz since the antenna elements in the centre of the station array have less physical area available than their maximum antenna aperture [Nijboer, 2009]. The calculated sensitivity values for the HBA at 140 MHz agree very well with observed values for the range 120 – 167 MHz that show little variation.

4.6.2 Number of sources per beam for self-calibration of ionosphere and beam shape

The interferometer sensitivity for different stations and frequencies is given in table 4.3 and can be rescaled to an equivalent $3\sigma_{\text{rms}}$ flux at 1.4 GHz to find the number of sources using table 4.2. In addition to the sensitivity in 10 s needed for rapidly varying ionosphere induced phase corrections we can use integrations up to 10^3 s and reach 10 times more sensitivity to find more sources that can be used for calibration of the beam shape and global refraction over the beam that varies more slowly.

The station main beam can be approximated reasonable well by a Gaussian profile $\exp(-r^2 / 2\sigma^2)$ that has levels of 0.78, 0.61, 0.37 and 0.14 at radii of 0.71σ , σ , 1.41σ and 2σ respectively. We have a central area of $\pi\sigma^2$ with average level ~ 0.78 and a next annulus with the same area and average intensity ~ 0.49 , that together cover about $\frac{1}{2}$ the area of the Gaussian profile with cut-off at 2σ . The actual station main beam has a first null at $\sim 2.4\sigma$, but the annulus between 1.4σ and 2.4σ covers only $\frac{1}{4}$ of the sensitivity weighted area.

In table 4.4 we convert the interferometer noise to an equivalent 1.4 GHz flux to estimate the number of sources per beam using the sources count for 1.4 GHz. The table clearly shows the effects of the varying spectral index with flux level in the transformation to equivalent 1.4 GHz flux and the effect of decreasing steepness of the cumulative source count. The result is a much slower increase in number of sources with increasing integration time at 140 MHz than at 70 MHz and 35 MHz. Some ratios in the table make larger steps than expected due to coarse steps of 0.1 in spectral index which introduce a factor $10^{0.1} = 1.26$.

An interesting aspect shown for 10 s sensitivity at 140 MHz is that the number of sources per beam is almost constant since we are in the integrated source count regime with index -1 where a larger telescope has more sensitivity that compen-

sates the loss of beam area. At higher sensitivity the number of sources per beam even decreases with station size. At low band frequencies the European stations are twice as sensitive as the Dutch ones but have a different distribution of the antenna elements over the station aperture leading to a different beam area.

Table 4.4. Number of calibration sources per central beam area with SNR > 3 for various LOFAR stations

Frequency	MHz	35			70		140	
Station type		LBA	LBA _E	LBA _S	LBA _E	HBA _C	HBA _R	HBA _E
Beam area *	deg ²	34.2	48.6	54.8	12.2	18.5	9.1	4.64
noise**	Jy	7.28	3.56	3.51	1.71	0.155	.0773	.0386
ΔS in 10 s ***	mJy	1464	716	1229	599	94.6	47.1	23.5
Sources/beam		1.1	5	2.3	1.43	36	35	36
ΔS in 10 ² s ***	mJy	463	226	389	189	33.6	16.7	9.36
Sources/beam		7.1	30	15	9.5	101	96	78
ΔS in 10 ³ s ***	mJy	146	71.6	123	59.9	11.9	5.93	3.73
Sources/beam		38	124	77	37	257	221	163

* central beam area spanned by $\pi\sigma^2$ with $\sigma = 0.425 \theta_{1/2}$ for Zenith direction.

** Noise per interferometer in 10 s at the different frequencies and bandwidths of table 4.3.

*** Flux level ΔS of sources at 1.4 GHz to give SNR = 3 including factor 0.78^{-1} by reduced average sensitivity over central beam area.

To model a delay screen that describes the large-scale TID sufficiently accurately we need at least one self-calibration source per 8 deg^2 with SNR > 3 per interferometer per 10 s, as will be analysed in subsection 4.8.4. However one source is not enough for an interpolation scheme, at least 3 are required that span a plane and allow linear interpolation in two directions. With 4 sources curvature in one direction can be handled as well, and curvature in two directions needs at least 5 sources.

At 140 MHz we find for the core stations ~16 sources per patch of 9 deg^2 at the centre of the beam, the remote stations have ~32 sources per patch but the European stations have effectively only ~16 sources per patch since only $\frac{1}{4}$ of the sources is smaller than $0.7''$ and not resolved on baselines with core stations. Apparently we can sample the delay screen much finer and even correct for small-scale disturbances if a sufficient number of independent baselines is indeed available for each station to allow such a solution in principle.

An important property of the integrated source count formula with slope index -1, which is valid for the relevant range of calibration source fluxes, can be summarized as follows.

Requiring on average 5 sources per central beam area that have $\text{SNR} > 3$ gives

- One source with $\text{SNR} \sim 15$,
- Two sources with $15 > \text{SNR} > 5$
- Two sources with $5 > \text{SNR} > 3$

The annulus around the central beam area has an average sensitivity that is a factor 0.63 lower, which means that we find only 3 sources with $\text{SNR} > 3$, i.e. one source with $\text{SNR} \sim 10$ one with $10 > \text{SNR} > 5$ and one with $5 > \text{SNR} > 3$. The total number of sources that could span a delay screen for $\frac{3}{4}$ of the weighted beam area with beam sensitivity > 0.37 is therefore 8 instead of 5. This allows to decrease the number of sources in the central beam area and we require only 3 sources in the central area providing 5 sources over the beam down to 0.37, which supports a 2-D curved delays screen model over $\frac{3}{4}$ of the sensitivity weighted station beam area.

The Dutch LBA stations need at 70 MHz ~ 20 s integration time to find 3 sources in the central area of the station beam that extends over 7 patches. The delay screen also covers the first annulus with 7 more patches, which shows that the sensitivity is not enough to provide sufficiently dense sampling. The main reason is that the station is too sparse as can be seen in the last row of table 4.1. The other reason is that we are in a regime of the integrated source count that has slope index -1.4, which means a rapid decrease of sources per beam for a small decrease in sensitivity.

The European LBA stations need at 70 MHz sources $< 1.5''$, which is $\sim 1/3$ of the population so a longer integration time of even ~ 100 s is needed to provide 3 sources in the central area of the beam. Including the first annulus a total of 3 patches are covered by the delay screen which is about enough for full beam calibration of a European station, but the sampling rate might be too low for accurate self-calibration.

The European LBA stations need at 35 MHz sources $< 3''$ which is still about $1/3$ of the population and need ~ 20 s integration time to provide 3 source in the central area of the station beam. The Dutch stations have a slightly smaller beam area, a factor lower sensitivity but no size constraint and need ~ 25 s to provide 5 sources over the beam. The delay screen over the station beam covers ~ 8 patches that provide more fine structure that cannot be corrected accurately by a delay screen that supports only second order interpolation for two directions.

The main result is that a larger interpolation error will be obtained and also a larger contribution by Kolmogorov turbulence is to be expected for interpolated positions further away from the reference ones.

4.6.3 Improving the spatial sampling for the delay screen

The previous subsection has shown that the LBA stations can observe sufficient sources to span a delay screen with 5 self-calibration sources over the beam using 20% relative bandwidth. However, the required sampling is 20 - 100 s, which might be adequate in good ionosphere conditions but is not sufficient in general. Moreover, the sampling of self-calibration sources is not dense enough causing larger phase errors for sources in between the reference ones. Additional tricks are needed and 3 possible ones will be discussed.

At 70 MHz three additional LBA station beams can be formed and positioned in a heavily overlapping configuration. For the European stations, with their factor two narrower beam w.r.t. the Dutch stations, we could increase the sensitivity of the piercing points on the sky and improve the match to the TID scale size. Alternatively the station could taper down the sparse outer rings, which widens the beam significantly and reduces sensitivity only slightly. More sources are then observed to span a TEC screen over the station beam.

Operating only with the Dutch stations that have a wider beam, the same approach could be used to improve the sensitivity in the annulus and increase the source density in the delay screen over the beam centred between the three additional beams.

At 35 MHz there are even six beams that could surround the central beam of Remote and European stations that have comparable beam width and provide additional sensitivity to fit in principle a curved delay screen that extends over the full size of the central station beam, even including the first side lobe.

4.6.4 Summary and conclusions for system use

The results of section 4.6 can be summarized as follows:

- With 20% relative bandwidth and 10 s integration time 16 sources per patch of 8 deg^2 could be observed by a HBA core station and 32 sources per patch with a remote station giving $\text{SNR} > 3$ in the central beam area. Such a source density is adequate to describe not only a large-scale TID profile but also the fine scale structure.
- Integrating over 100 s the increases the sensitivity resulting in more sources per station beam and could improve the spatial accuracy of the

delay screen model. To avoid reduction of the temporal accuracy a tracking approach might be used as will be discussed in subsection 4.8.4.

- The quoted -small- numbers are averages where Poisson statistics causes large variation that could limit the Calibratability of specific fields.

Conclusions for system use are:

- Beam size and sensitivity of all LOFAR LBA stations are marginally adequate to allow self-calibration using 5 sources with $\text{SNR} > 3$ per station beam down to 0.37 of the peak value. Relative bandwidth of ~20% and integration times of 20-100 s are then required, which is just adequate to correct for the large-scale TID on baselines with a sufficiently low rate of change in the TEC induced delay.
- A delay screen model using 5 sources supports only 2nd order interpolation over $\frac{3}{4}$ of the sensitivity weighted beam area.
- The European HBA stations need the baselines to the Dutch stations with length < 600 km that do not resolve the $\frac{1}{4}$ fraction of sources $< 0.5''$ and have 14 sources per beam down to 0.37 of the peak sensitivity.
- To improve the sampling density by self-calibration sources, the LBA stations could use dense multi-beaming to provide a delay screen that describes the complete central beam that has a much larger extent than the TID patches.
- The number of sources per beam could be extended for the European LBA stations by applying a station taper that increases the beam with minor loss of sensitivity and is important for observing at frequencies higher than 50 MHz.

4.7 From interferometer phase to station based TEC screen values

The simplest ionosphere self-calibration models [Cotton, 2004], [Cohen, 2007], [Tol, 2009], [Intema, 2009] assume a thin phase screen that induces only a phase jump when a ray passes. Rays from telescopes to sky sources have puncture points on this phase screen. If it is known that the maximum phase difference between two puncture points is less than π the sign of the phase gradient can be determined unambiguously. This allows estimation of the phase as function of direction of each ray from each telescope by interpolation between phases derived from sources in only a few directions. A key parameter of the phase screen is its height above the telescopes, which is needed to convert differences in angle between the rays from each telescope to differences in screen position. Unfortunately when observed interferometer phases are decomposed into telescope phases, an independent and arbitrary offset is found for each source direction [Hamaker, 2000]. In principle these offsets could be chosen such that in first instance the station phases are zero

for all directions from an arbitrarily chosen reference telescope. In a next step the offset per direction needs to be modified such that the phase screen over the array also gives the proper phase values above the reference telescope. Although any remaining arbitrary offset drops when interferometry corrections are made for the reference sources, the interpolation process should be such that this is also true for interpolated directions.

Either interferometer phase originates from physical delay differences between the two stations or from differences in phase corrections applied by the processing in each signal chain or by phase corrections applied to the interferometer data. It is therefore proposed to derive delay differences from observed interferometer phase measurements over a sufficiently large frequency band and work with a delay slab instead of a phase screen to eliminate π ambiguities. The excess delay along the path towards each station is determined by the TEC in Zenith direction and causes an actual delay that is proportional to the secant of the actual Zenith angle of each source and the location of each station projected on the delay slab. In addition to a height parameter, a thickness parameter is also required for the delay slab model.

Typical TEC values of ~ 20 TECU valid for LOFAR correspond to an excess pathlength of ~ 800 m at 100 MHz (subsection 4.1.1), which is small compared to some average ionosphere thickness of order 300 km (intro section 4.2). A thin excess pathlength screen that only describes the gradients in horizontal direction can be determined from phase measurements on a single source by a set of interferometers. A sufficient range of baselines and frequencies needs to be spanned such that interferometer refraction can be eliminated while a sufficient number of baselines near zero length avoids phase ambiguities. Even sources with different directions can be included in the delay screen model but this requires not only an additional parameter for the height of the screen but also a proper procedure to handle station offsets per direction.

Although these offsets are arbitrary from the mathematical point of view [Hamaker, 2000], they have a physical counterpart in the thickness of the curved homogeneous slab, which eliminates them in a physics based model.

The core area of LOFAR has an extent that is comparable to the ionosphere area covered by a station beam, allowing solving for a delay screen that extends over a number of overlapping telescope beams. With such a properly defined ionosphere reference patch for the core of the array also the differential delay distribution over patches above remote stations can be determined unambiguously from wide band interferometer measurements.

4.7.1 From interferometer phase to delay, TEC and phase unwrapping requirements

An interferometer array observes phase differences relative to the phase of one of the telescopes as a function of time at a range of frequencies. Path delay is the physical property in propagation of a wavefront through atmosphere, antennas and receivers before the spatial correlation is established. Therefore all observed interferometer phases have phase delay differences as physical cause except for some types of instrumental cross talk and for improperly applied phase corrections. We can therefore express the observed phase $\varphi_{\text{obs}}(t, \nu)$ as a function of frequency ν [GHz] and as a function of differential $\Delta N_{\text{TEC}}(t)$ [TECU], differential delay $\Delta\tau(t)$ [ns] and instrumental offset phase $\varphi_{\text{instr}}(t)$:

$$\varphi_{\text{obs}}(t, \nu) = 2\pi \nu (1.34 \Delta N_{\text{TEC}}(t) \nu^{-2} + \Delta\tau_{\text{clock}}(t) + \Delta\tau_{\text{instr}}) + \varphi_{\text{instr}}(t) \quad (4.29)$$

For a narrow band observation it is not possible to separate observed delay accurately in a proper TEC contribution and a true time delay contribution. We can separate the delay between the two stations in a fixed instrumental differential offset $\Delta\tau_{\text{instr}}$ and two variable components, one determined by the difference in Total Electron Content $\Delta N_{\text{TEC}}(t)$ along the ray paths in TECU and one for the difference in clock time $\Delta\tau_{\text{clock}}(t)$.

In practice also instrumental delay can show dispersion due to differences in the pass band filter characteristics of the receivers, which can however be removed by an initial pass band correction as first step in the processing. Fortunately, these instrumental terms do not change with time, and clock differences vary on time scale $> 10^3$ s while TEC changes at 10 s time scales, allowing a proper separation in principle.

With 5 sources per beam that all have $\text{SNR} > 3$ we can determine delay and phase offset per interferometer. The total observed bandwidth allows the strongest self-cal source to reach $\text{SNR} > 15$ assuming that we are in the regime of the integrated source count with index -1. By separating this total observed bandwidth into 9 spectral bins each will have $\text{SNR} > 5$ giving ~ 0.2 rad rms phase noise per bin for the strongest object, which limits phase uncertainty and allows unambiguous phase unwrapping over the frequency range to derive the delay. This delay can also be used as reference for the weaker self-cal sources to resolve potential phase ambiguities in the delay estimation for these weaker sources.

The maximum clock difference between two stations is < 10 ns giving a phase change of < 0.3 rad over a spectral bin < 5 MHz (applicable at 140 MHz observing frequency). A TID induced maximum TEC difference < 0.2 TECU over distances > 50 km gives on longer baselines a maximum delay difference < 220 ns at 35 MHz, which results in a phase change < 1 rad for a spectral bin separation < 0.7 MHz (as

is the case at 35 MHz) and allows proper phase unwrapping and also good TEC estimation.

However, for European baselines we have to deal with a wedge gradient that could be ~ 1 TECU over 600 km from the LOFAR core which could give a phase change up to ~ 4.8 rad over a spectral bin of 0.7 MHz (as for 35 MHz observing frequency), which requires a smaller bin width for proper phase unwrapping.

In practice we derive initial calibration parameters from a snapshot set of visibility data using integration times 10 – 100 s which, according to table 4.4, contains more than three sources per beam with $\text{SNR} > 3$ for all LOFAR stations at all LOFAR frequencies. So there will be a strongest source allowing more than 9 spectral bins which is adequate to establish initial atmospheric TEC, instrumental delay and phase offset parameters for all LOFAR stations from all interferometers, except from interferometers between distant European stations.

4.7.2 Decomposing Interferometer delay and TEC into station based delay and TEC

Since all interferometer corrections originate per station we can decompose them into station-based parameters. This might not be true for some crosstalk signals as discussed in the previous subsection, but can be ignored in practice. However, decomposing TEC, delay and phase introduces an arbitrary common offset per source direction in TEC, in delay as well as in the residual phase, which drops out when the difference between two station corrections is used to correct an interferometer. We can therefore subtract from each decomposed station TEC and each decomposed station delay and each decomposed station phase an arbitrary equal value for all stations. In practice we choose for this value the average over a subset of stations for which the variation over time is known to be small, which allows following individual station values that have large variation over time that might indicate malfunction of a particular station.

Although each set of station delays for a particular direction has its own arbitrary offset, the differences between delays for each station is well defined by the TEC differences between the different directions. Also there could be an instrumental delay offset between directions if an improper phase reference centre for a station is used as discussed in subsection 3.6.1. This delay offset can in principle be measured with a holographic measurement setup, but is not observable in a synthesis observation. Correction for such a potential station position error has to be applied in a preceding calibration step together with a nominal pass band correction and a correction for nominal refraction. In a second step initial corrections, derived from the strongest source, are applied to the interferometer data, which eliminates all direction independent instrumental terms such as clock and phase zero. Unfortunately, also some TEC is eliminated by this step since the separation between

TEC, delay and phase for the reference source direction is not perfect due to limited bandwidth. In subsequent steps differential corrections are determined for all other sources in the beam that have $\text{SNR} > 3$.

These differential phase corrections then contain pure excess TEC by the ionosphere, while the relative amplitudes are defined by the beam shape profile. As a result the differential phase delays of a set interferometers can be decomposed into a station based TEC contribution for each source direction. Unfortunately an arbitrary offset per set of station values is present for each direction, which needs to be eliminated to define a proper TEC screen for the synthesis array.

The impact of the arbitrary offset term for each source direction depends on how flat the screen is. Figure 4.11 shows a cartoon with the geometry of the delay wedge that consists of a planar slab with a thin wedge on top for two typical LOFAR situations.

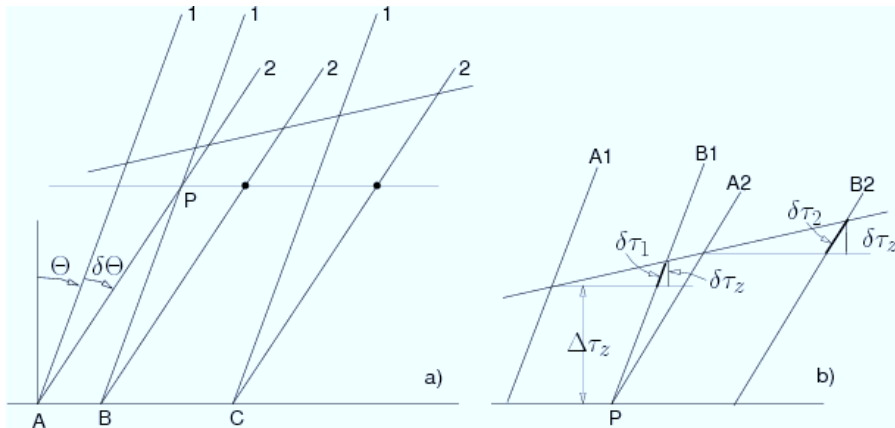


figure 4.11. Interferometer excess pathlength differences for different zenith directions θ and $(\theta + \delta\theta)$.

(a) Telescopes A, B & C observe objects 1 & 2 through a homogeneous wedge with an excess delay in zenith direction with a constant contribution $\Delta\tau_z$ and a variable term $\delta\tau_z$ that is proportional to the TEC gradient and to the distance between the stations

(b) Blow-up of the area around piercing point P of rays A2 and B1. Interferometer AB observes object 1 with differential excess delay $\delta\tau_1 \sec(\theta)$ but object 2 with delay $\delta\tau_2 \sec(\theta + \delta\theta)$, since the equal ray parts for each source direction cancel in the flat layer below the small wedges as indicated by the parallelograms. The small wedges between stations A and B for directions 1 and 2 give the same position shift as the large wedge between A and C in (a) although a larger differential time delay is involved on longer baselines.

As explained in caption b) of figure 4.11 the delay in the uniform bottom part of the wedge has no influence on the observed phase of the interferometers since the rays have equal pathlength in the parallelograms. This is perfectly true if this bottom part is indeed a homogeneous planar slab, and is even true if vertical stratification would be present as discussed in subsection 4.1.3. However, a curved slab will give spherical refraction as discussed in subsection 4.1.6. This means that the arbitrary offset in the station solutions is no longer arbitrary, but is in fact the slanted thickness of the uniform slab below the wedge, which causes differential spherical refraction over the station beam.

4.7.3 Large-scale refraction effects

The planar slab in figure 4.11 with thickness $\Delta\tau_z$ is indeed curved and produces spherical refraction as discussed in subsection 4.1.6. Applying a nominal spherical refraction correction to the interferometer data that is valid for the centre of a field is a first correction step to the data as discussed in the previous subsection, but we are still left with differential spherical refraction and with wedge refraction. Nominal refraction correction is according to (4.16) about 50" (at 100 MHz) at zenith angle $\theta = 45^\circ$ and 62" at 48° , while typical wedge refraction at 45° elevation is ~24" according to (4.13). A nominal refraction correction is applied to the visibility data but has a typical 10% error, which amount to 5" at 45° elevation and 6.2" at 48° at half power of the station beam. Self-calibration with a defined nominal position for the strongest source defines the position of the whole field and includes correction for the actual residual spherical refraction of the reference source and for the wedge term contribution of the reference source. Differential spherical refraction correction over the field can in principle be corrected by rescaling the image after Fourier transformation relative to the position of the strongest source in the field. In practice the sum of the two differential refraction corrections that scale both with frequency squared leads to station based direction dependent TEC values derived from additional reference sources in the station beam. Although spherical refraction and wedge refraction are both proportional to $\sec^2\theta$ that can be simply corrected for, we need model fitting to solve for a TEC screen with zenith angle values that need the $\tan(\theta)$ factor to describe the differential spherical refraction contribution.

4.7.4 Differential delay screen corrections using a peeling approach

We assume a 'peeling' approach [Noordam, 2004; Tol, 2007] for the self-calibration. In this approach, a first set of calibration parameters is estimated for the strongest source in the field using a proper nominal source model for that source. The visibilities are then corrected and the nominal source model for this strongest source is subtracted. There are still more sources in the field strong enough to derive self-calibration parameters, which provide relative corrections for their respective sky

directions. These differential corrections are obtained by using a nominal position for each source and therefore correct not only for a shift by local TEC gradient of a TID, but also for the difference between the actual refraction and the correction applied for the actual phase deviations of the strongest reference source.

Using the interferometer phases for a number of additional sources the decomposed differential station phases only include differential TEC and could be used to fit a TEC screen model defining the differences relative to the direction of the strongest source, for which corrections are already applied. In such a model fit each set of station TEC values for a specific direction θ needs an appropriate $\sec^2(\theta)$ correction to find the nominal zenith value for the model. Although the station TEC values of the first additional source could be arbitrarily normalized to have zero average (over all stations), each next set needs an additional offset parameter to obtain a smooth screen without jumps for piercing point for a specific source direction.

When sources at nominal positions are subtracted using phases derived from this smooth fit TEC screen, also the differential refractions by curved slab and wedge are properly included, while the residual visibilities are not corrected. An image of these residual visibilities still needs correction for the TEC screen, for instance by using a convolutional correction like for the W-term that is also the difference of station contributions.

When stations have a smaller separation than the linear scale size of the station beam at the height of the atmosphere phase screens, then spatial station sampling is turned into additional angular TEC screen sampling. This means that less sources per beam need to be solved for stations in clusters where they are closer together than the size of the atmosphere structures. This relaxes the minimum sensitivity for full delay screen calibration for stations in clusters, with average separation less than 10 km.

4.7.5 Accuracy of station based phase delays

The peeling approach introduced in the previous subsection starts with the strongest source and solves for a complex gain correction per station. The corrections have an accuracy not only determined by the noise per interferometer, but also suffer from contamination by all other sources in the sky. To reduce this contamination an iterative procedure is followed where the solution of stronger sources is improved when the next weaker source is solved for. It has been demonstrated that this procedure indeed works and reaches the nominal noise level when all sources are solved and an appropriate correction is made to eliminate bias effects [Tol, 2007]. Unfortunately, only sources stronger than 3 times the thermal noise in an interferometer can be solved this way, and we need to investigate the effect of weaker sources that are not solved for.

In the following paragraphs we give a first-order estimate of the expected thermal noise in a solved parameter and compare that with the additional noise introduced by unsolved sources that are still present in the visibility data.

When there are M sources per beam with $\text{SNR} > 3$ there are also M additional sources with $3 > \text{SNR} > 1.5$ (assuming index -1 for the integrated source count). The additional sources have an average flux of about $\sim 2.2 \Delta S$ where ΔS is the interferometer sensitivity as discussed in section 4.6. For a first-order estimate of the disturbance by these sources we could aggregate the M additional sources to a single source assuming equal strength and uniform distribution of the phases giving an equivalent strength $2.2 M^{1/2} \Delta S$. Since the phase of each constituting source is different for each interferometer also the phase of the aggregate source is different for each interferometer and we assume a uniform distribution over 2π for the phase of the aggregate source over a set of interferometers. The decomposition in station delays effectively averages for each station over N_{ind} independent interferometers to at most $N_{\text{st}} - 1$ other stations and introduces a phase error equal to a relative amplitude error in the decomposed parameters for a calibration source with flux S_{cal} given by

$$\delta\varphi_{M \text{ sources}} \sim 2.2 (M / N_{\text{ind}})^{1/2} \Delta S / S_{\text{cal}} \quad (4.30)$$

We could include the next set of $2M$ sources in the bin $1.5 > \text{SNR} > 0.75$ that have half the average flux (if still on the same -1 slope of the integrated source count) and find a separate rms contribution equal to $0.7 \delta\varphi_{M \text{ sources}}$. This procedure could be extended and adding all contributions in squared sense gives a total factor $(1 + \frac{1}{2} + \frac{1}{4} + \dots)^{1/2} = 1.41$ in (4.30) for all sources weaker than the M sources that are included in an iterative peeling solution.

We concentrate on the weakest calibration source and we take the situation with 5 sources per beam, which means roughly one source in the centre of the beam and the other four at half power. The distance from a self-cal source to disturbing weaker sources is then order $1/4^{\text{th}}$ of the width of a station beam. The peeling process puts the fringe tracking centre at each self-cal source and all disturbing sources get bandwidth attenuation as if at half power of a station that has twice the diameter. According to (3.32) in subsection 3.2.3 the disturbing sources are 1.7% degraded at half power distance of a station of $\sim 2 \times 40$ m diameter for a relative bandwidth of 1.3 % at baselines of 1 km. This degradation increases quadratic to 85% loss for a relative bandwidth of 9%. For larger bandwidth, the main lobe of the sinc function becomes narrower than the station beam and the attenuation is described by the side lobes of the sinc. These side lobes, with amplitudes smaller than 0.13, reduce the contribution to phase error (4.30) by most sources in the beam to $\sim 10\%$. Bandwidth decorrelation works best for sources with a position offset in the direction of a baseline and less for other directions. Since our actual relative bandwidth is $\sim 20\%$ instead of 9% we assume our aggregate source reduced in intensity by a factor ~ 10

and including the factor 1.41 discussed above we find a phase error in the weakest self-calibrator with $S_{\text{cal}} = 3\Delta S$ given by

$$\delta\varphi_{\text{all sources}} \sim 0.1 M^{1/2} N_{\text{ind}}^{-1/2} \quad (4.31)$$

We still have a thermal noise contribution for the weakest source with $\text{SNR} = 3$ given by

$$\delta\varphi_{\text{thermal}} = 0.33 N_{\text{ind}}^{-1/2} \quad (4.32)$$

Therefore, the accuracy of the delay and amplitude solution for the weakest of $M < 10$ self-calibration sources with $\text{SNR} > 3$ is only dominated by thermal noise when baselines longer than 1 km and relative bandwidth larger than 9% are used.

4.7.6 TEC screen construction by renormalization of station based direction dependent TEC

The TEC screen over a station beam in the LOFAR core area is sampled by a number of interferometers that have shorter baselines than the width of the station beam at the height of the delay screen. This means that the number of piercing points per station beam is increased as indicated by the cartoon in figure 4.12 and enables a TEC screen reconstruction with less than 5 sources with $\text{SNR} > 3$ per station beam for the core area.

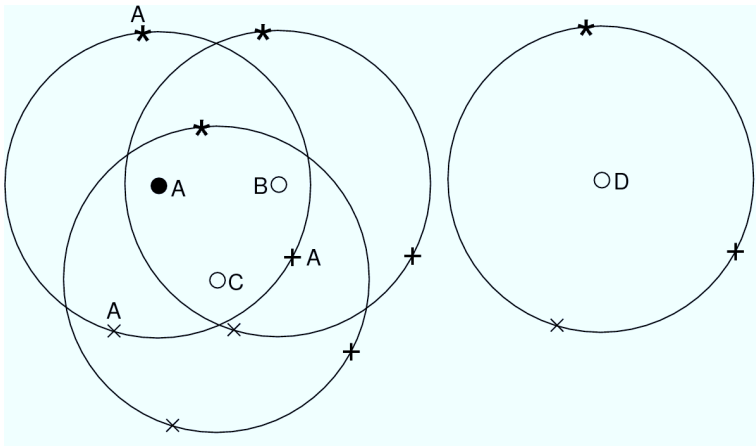


Figure 4.12. Piercing points in a delay screen spanned by 4 sources o, x, + and * in an array with 4 stations A, B, C and D. Reference source o is at the centre of the station beam and the three other sources are approximately at half power.

The cartoon gives a simplified 2-D picture and shows dense sampling of the TEC screen by stations in the core area, while a remote station only has coarse sampling by different sources. Stations A, B and C are so close together and their beams so wide that rays to different objects have piercing points close together, such that the same piece of the TEC screen is sampled by different interferometers. Station D is so far away that the TEC screen over its beam is not sampled by other telescopes, requiring more than 3 sources with $\text{SNR} > 3$ per beam to solve for curvature in the delay screen over that beam. For a typical beam width $\sim 6^\circ$ for the core stations and assumed TEC screen height of ~ 215 km a screen area with 22 km diameter is spanned covering piercing points from stations in the core and remote stations out to 22 km.

Since we fully corrected all visibilities for the station values of the strongest reference source, in our case at the centre of the beam, all station TEC values for this reference direction are zero. The differential station TEC as derived from interferometer data of another source contains not only the true TEC value that is different for each station, but also an arbitrary offset common to all stations. This arbitrary common offset needs some renormalization, which can be derived from comparison between piercing points that are close together but originate from different source directions. In fact there are two types of renormalization, (i) by adding a value to all station TEC values for a certain direction, (ii) by correcting each station for the zero TEC value of the strongest reference source. The first renormalization type corrects for the initial assumption that the differential TEC values of all directions are zero for the arbitrarily chosen reference station, in our case station A. The second renormalization type corrects for the assumption that TEC in the reference direction of each station is absorbed in instrumental delay, leaving no contribution to the local TEC screen.

A simple renormalization approach halves the difference between two close piercing points from two different source directions by adjusting the common offsets by equal but opposite amounts, i.e. type (i). If one of the piercing points is the reference source of a station, we correct only the station value, i.e. type (ii). This process has to be repeated for at least one piercing point of each source direction. The whole cycle can be repeated a couple of times and we expect for the same pair of piercing points smaller corrections in subsequent steps.

If a station value of the reference source needs to be changed we need to move an applied initial visibility correction to the TEC screen correction. We need therefore two corrections (i) change the TEC correction of the station that was applied to all visibilities before peeling of the other sources, (ii) change the TEC screen correction of the station for all other source directions (that are relative to the correction of the reference source).

The offset renormalization needs to include the secant of the zenith angle for each source direction. The maximum change ΔTEC over 11 km by a TID is ~ 0.07 TECU and the change δTEC by renormalization that includes the difference in $\cos(\theta)$ is given by

$$\delta\text{TEC} = \Delta\theta \sin(\theta) \Delta\text{TEC} \quad (4.33)$$

So, the renormalization difference between two close piercing points for directions with difference $\Delta\theta \sim 3^\circ$ at $\theta \sim 45^\circ$ is only 0.0026 TECU or a phase of 0.22 rad at 100 MHz.

Further discussion of a detailed procedure is outside the scope of this summary discussion, but the three principal issues of arbitrary normalization per source direction, initial zero TEC screen values for the strongest reference source and effective TEC screen thickness involving the secant of the actual zenith angle have been analysed.

The details of an actual iterative process are not critical since interferometer corrections are always derived from station differences, which eliminates any arbitrarily introduced offset in the delay screen. The main purpose of the procedure is to obtain a delay screen with realistic derivatives and no sudden jumps, such that a simple interpolation process is sufficient to find the station corrections for any other source direction within the station beam.

We outlined a procedure that addressed the arbitrary phase offset in a station solution, which is in effect only one term in the unitary matrix that describes the full polarization characteristics [Hamaker, 2000]. Recently a solution for this more general problem has been proposed and demonstrated [Yatawatta, 2012a].

4.7.7 Summary and conclusions for system design

Our analysis of station based ionosphere TEC modelling can be summarized as follows:

- It is assumed that nominal corrections are made for nominal ionosphere refraction and instrumental pass band to all visibilities.
- Using the strongest calibration source in the beam we can unwrap phase rotation over the wide pass band and solve for TEC, clock delay and residual phase for each interferometer.
- A potential residual phase term can be attributed to the signal chain for instance by incomplete pass band correction.
- Baselines shorter than about 1 km between stations in the core array, should be excluded from such station based solutions. A better defined limit requires a more detailed analysis.

- After decomposition into station based terms, corrections to all visibilities should be made that correct for differences in instrumental delay and phase, station clocks, and ionosphere TEC for that reference source.
- We get proper correction for the reference direction irrespective of improper separation between delay and TEC as a consequence of limited bandwidth.
- The peeling approach allows finding TEC in different directions for each station relative to TEC for the strongest source in the beam that is assumed to be zero.
- A TEC screen can be constructed that combines station TEC for different source directions using a renormalization procedure for station based TEC per source direction that includes inclination effects by the thickness of the disturbances over the screen.
- Also renormalizations of the station corrections for the reference direction are required to restore TEC values that fit in the TEC screen over the array.
- Although we do not obtain the true screen since the true thickness is still arbitrary, proper differential phase correction for each station in each direction can be obtained.
- An estimate for the true thickness of the curved slab might be obtained from fitting differential refraction over the FoV.

Conclusions for system design are:

- We have shown that the solutions of the weakest reference sources are thermal noise dominated if less than 10 sources with $\text{SNR} > 3$ per interferometer are solved while all weaker sources are ignored, provided that the used baselines are longer than 1 km and the relative bandwidth is larger than 20%.
- A TEC screen over the synthesis array can be constructed that uses 5 self-calibration sources per station beam and allows noise dominated phase corrections for all other sources in the beam that are too weak for appropriate individual self-calibration.
- This requirement could be relaxed for the core area, since we need less sources per beam of 6° width to define a delay screen with many piercing points that covers an area of 22 km diameter around the projection of the core on the ionosphere delay screen.

4.8 Simplified polynomial interpolation model for the delay screen

One method of defining coefficients in a polynomial model is based on generating moments where the data points get a weight relative to their distance from a reference position and relative to their SNR. The moments that together provide phases for all positions are dominated by the strongest source in the field with the highest SNR. However, at the position of each self-cal source we do not get back the exact phase of that source but a phase that includes contributions by the other self-cal sources. We therefore look into an interpolating scheme that maintains the values of the reference sources at the reference positions. One example of such a scheme, the Lagrange interpolation, will be analysed to reveal some characteristic features and is the subject of this section.

4.8.1 Lagrange interpolation

To demonstrate the effects the simplest one-dimensional second order Lagrange interpolation is used with interpolated value $y(x)$ given by expression:

$$y(x) = y_0 L_0(x) + y_1 L_1(x) + y_2 L_2(x) \quad (4.34)$$

The values y_i are the actual ones at position x_i and the Lagrange polynomials are given by

$$L_0(x) = (x - x_1)(x_0 - x_1)^{-1}(x - x_2)(x_0 - x_2)^{-1} \quad (4.35a)$$

$$L_1(x) = (x - x_0)(x_1 - x_0)^{-1}(x - x_2)(x_1 - x_2)^{-1} \quad (4.35b)$$

$$L_2(x) = (x - x_0)(x_2 - x_1)^{-1}(x - x_1)(x_2 - x_1)^{-1} \quad (4.35c)$$

A further simplification uses $x_0 = -1$, $x_1 = 0$ and $x_2 = +1$ and allows evaluation of the polynomials for a representative range of intermediate values in table 4.5.

Table 4.5. Lagrange coefficients for some values of the argument

	x	-5/4	-1	-3/4	-1/2	0	1/3	2/3	+1	4/3
L_0	$(x-1)x/2$	45/32	1	21/32	12/32	0	-1/9	-1/9	0	2/9
L_1	$(x+1)(1-x)$	-18/32	0	14/32	24/32	1	8/9	5/9	0	-7/9
L_2	$(x+1)x/2$	-5/32	0	-3/32	-4/32	0	2/9	5/9	1	14/9

The polynomial functions in the table represent a normalized weight, i.e. their sum is 1 for all values of the argument, while $L_1(x) = L_1(-x)$ and $L_0(x) = L_2(-x)$ form an symmetric pair.

The values of the polynomials in the table clearly demonstrate that at the location of the reference points only the reference value is used and that reference points that are furthest away from a required position get the lowest weight. Beyond 2/3 of the sampling distance from the central source starts the nearest source to dominate. This result can be generalized as expressed by the following statements:

- Noise in a Lagrange interpolated value is dominated by the SNR of the closest reference point
- Noise in reference points that are further away play a minor role.

A different issue is the interpolation error and will be analysed in the next subsection.

4.8.2 Accuracy of 2nd order Lagrange interpolation for a TID sine wave model

We are now in a position to compare the results of a Lagrange interpolation with for instance a true sinusoidal shaped delay screen valid for a TID. We analyse a representative situation where the three reference points are separated by 45°, but cover different segments of the sine curve. One segment covering 0°-90° has a long straight and a short curved part that are not evenly sampled, which is a typical worst case scenario for 2nd order interpolation. The other segment covers 30°-120° that has a short straight part and a long curved part

Table 4.6. Second order Lagrange interpolation errors for sine segments

X	-5/4	-1	-3/4	-1/2	0	1/3	2/3	+1	4/3
φ	-11.25	0	11.25	22.5	45	60	75	90	105
$\sin(\varphi)$	-0.195	0	0.195	0.383	0.707	0.866	0.966	1.000	0.966
$L(\varphi)$	-0.241	0	0.215	0.405	0.707	0.851	0.948	1.000	1.001
$L - \sin$	-0.046	0	0.020	0.022	0	-0.014	0.018	0	0.035
φ	18.75	30	41.25	52.5	75	90	105	120	135
$\sin(\varphi)$	0.321	0.500	0.659	0.793	0.966	1.000	0.966	0.866	0.707
$L(\varphi)$	0.295	0.500	0.669	0.804	0.966	0.995	0.962	0.866	0.706
$L - \sin$	-0.026	0	0.010	0.011	0	-0.005	-0.004	0	0.001

In table 4.6 we subdivide the two intervals according to the increments chosen in table 4.5 for which the polynomial coefficients are evaluated. We get typical interpolation errors of 0.02 in the worst case scenario and errors of 0.04 for small extrapolation, but these could all be halved if the sampling is chosen differently as demonstrated by the second example.

In practice our choice is not free but determined by the actual positions of sources and TID, so we need to take the worst case results from table 4.6 as representative. Comparable results are obtained by retaining the linear term in a series expansion for the sine and retaining the quadratic term in a cosine expansion. Apparently a sampling range of 90° gives $\sim 100^\circ$ useful range which corresponds to ~ 25 km delay screen extent for a TID wavelength of 90 km. At an assumed delay screen height of ~ 215 km, we need 5 reference sources at separations of ~ 11 km that span a total area of $\sim 6^\circ$ diameter, to allow useful corrections for an area with $\sim 6.6^\circ$ diameter.

For a typical TID wave amplitude of 0.1 TECU we get interpolation errors of 0.002 TECU that are smaller than the local deviations of 0.005 TECU in the sine wave pattern as shown by figure 4.7. For TIDs with longer wavelength than 90 km the second order interpolation becomes even more accurate.

We assume that a two-dimensional interpolation has the same properties as the one-dimensional case and we conclude

- Second order Lagrange interpolation using 5 reference sources that span an area with $\sim 6^\circ$ diameter gives TID interpolation errors over an area with 10% larger diameter that are smaller than ionosphere induced disturbances.
- Lagrange interpolation gives exact correction at the reference objects and the smallest errors close to the reference points.

4.8.3 Delay screen accuracy limitations by Kolmogorov Turbulence

In addition to the large-scale structure in the delay screen by TIDs there is finer scale structure that we may characterize by Kolmogorov Turbulence. This finer scale structure cannot be described by interpolation based on sampling that is only adequate to describe larger scales and we need an estimate of the Kolmogorov Turbulence contribution.

We could consider correction for the large-scale TID effect, as discussed in the previous subsection as a form of tip-tilt correction for a local part of the delay screen as discussed in subsection 4.3.3.2. We therefore expect a rms phase noise relative to the nearest reference position in the delay screen as described by (4.21a). As shown in subsection 4.3.3.3 a large-scale delay screen that models half a TID wave of ~ 45 km extent has typically fine structure with maximum deviations from the sine

wave pattern up to 0.005 TECU that are ~ 15 km apart. Tip-tilt correction using effective sampling every ~ 11 km therefore removes already a part of the small-scale Kolmogorov variation and we can estimate the value for the residual phase deviation over an area around a reference point using (4.21a). This rms of the phase variation can for a given frequency be converted to an rms TEC variation over aperture A with diameter B using (4.3) at a reference frequency of 100 MHz and we find

$$\sigma_{\text{TEC } A} = 0.0047 (B / r_{100})^{5/6} \quad [\text{TECU}] \quad (4.36)$$

where r_{100} equals r_0 at the reference frequency. We convert (4.36) to the angular domain and work with the radius R_α of the aperture A and get

$$\sigma_{\text{TEC } A} = 0.0047 (R_\alpha / \alpha_{100})^{5/6} \quad [\text{TECU}] \quad (4.37)$$

where α_{100} corresponds to $\frac{1}{2} r_{100}$ at a height of 215 km assumed for the TEC screen. For situations where the phase screen is dominated by TIDs we have according to subsection 4.3.2.1 a typical value $r_{100} = 6$ km, which corresponds to $\alpha_{100} = 0.8^\circ$. Equation (4.21a) is valid out to a diameter of $3 r_0$ and is based on the phase structure function, which leads to a frequency dependent r_0 corresponding at 100 MHz to a maximum distance of 2.4° from a reference source. Equations (4.36) and (4.37) use a fixed r_{100} and frequency dependence of the phase, and are linear with wavelength according to (4.3). In practice the value of 2.4° also corresponds to the maximum sampling distance for which 2nd order interpolation for the shortest TIDs of ~ 90 km wavelength is useful. This couples the characteristic distance r_{100} to the physical scale for disturbances instead to observable phase at a given wavelength. We will ignore the implications of these aspects for our derivation of first-order estimates.

Subsection 4.3.3.2 has shown that under benign ionosphere conditions we find values $r_{100} > 15$ km or $\alpha_{100} > 2^\circ$ while subsection 4.3.3.3 indicate even much larger values in case correction for a TID pattern has been made.

Equations (4.21a), (4.36) and (4.37) describe the average rms over an aperture, which is the result of averaging the variance over the aperture. For a variance as function of radius $v = r^{2\beta}$ we get an average variance $(1+\beta)^{-1} r^{2\beta}$ which means that the maximum expected rms deviations at the rim of the area are a factor $(1+\beta)^{1/2}$ larger than given by (4.21a), (4.36) and (4.37). We simplify (4.37) by using unity exponent and find the rms TEC deviation over the rim at distance R_α by

$$\sigma_{\text{TEC } R} = 0.0064 R_\alpha / \alpha_{100} \quad [\text{TECU} / ^\circ] \quad (4.38)$$

Decreasing the sampling distance for a 2nd order Lagrange interpolation would improve the accuracy of the sine and cosine parts of the TID wave with 3rd and 4th order coefficients as follows from their series expansions. However, residual rms in

TEC by Kolmogorov turbulence has a power law distribution with exponent 0.83 and decreases less than linearly with the sampling distance.

We conclude

- The accuracy of interpolated delay screen values using 2nd order Lagrange interpolation between source directions is limited by Kolmogorov turbulence.
- Under typical conditions where short wavelength TIDs appear the residual TEC deviations are about 0.008 TECU/^o from the nearest reference source to at most 2.4^o from that reference point for $\alpha_{100} = 0.8^{\circ}$.
- In good ionosphere conditions we find a factor 3 lower rms deviations and a factor 3 larger extent as a result of increased α_{100} .

4.8.4 Matching station beam width and effective integration times

In subsection 4.6.2 we derived the number of sources per central beam area that have SNR > 3 per interferometer. We assumed a Gaussian beam profile $\exp(-r^2 / 2\sigma^2)$ and a beam area defined by $\pi\sigma^2$. The annulus with outer radius 1.41σ has the same area but lower average sensitivity and observes only 2 sources when the central area observes 3 sources with SNR > 3. A requirement of 3 sources with SNR > 3 in the central beam area therefore provides 5 sources that span a delay screen that allows 2nd order interpolation for all directions covered by $\frac{3}{4}$ of the sensitivity weighted station beam out to 37% of the peak sensitivity.

If the average separation between the sources is 3° we can interpolate medium scale TIDs with wavelength as short as ~90 km with an accuracy that is higher than residual TEC variation by small-scale Kolmogorov turbulence.

The remote HBA stations have sufficient sensitivity at 140 MHz, such that the total available processing bandwidth for station beam forming and array correlation could even be used to define 6 surrounding beams. These additional beams each have lower effective bandwidth but could by sparse distribution of spectral channels still span more than 20% relative bandwidth to allow effective peeling. The six surrounding beams cover the low level region of the main beam and the first side lobe and would extend the delay screen with more self-calibration sources at some 3° grid for adequate delay screen interpolation.

An initial analysis of relevant time constants that determine the allowed integration time for estimation of interferometer delays is given in subsection 4.3.2.4, and we now look a bit closer using details of the Lagrange interpolation process.

A TID propagating with a speed of ~150 m/s takes 40 s to travel 6 km, which corresponds to an angular distance of 1.6° at a screen height of 215 km. When the large-

scale TEC gradient over such a distance is removed, the Kolmogorov turbulence causes according to subsection 4.8.3 residual excess delay changes of ~ 0.0016 TECU/km (rms) leading to a change of 0.0096 TECU over 6 km giving a change $\delta\phi = 0.81$ rad at 100 MHz. After integration over an interval of 40 s we could then expect an amplitude decrease for which a first-order estimate is given by the factor $\text{sinc}(\delta\phi/2) \sim 0.973$. A much longer integration time of 100 s would then give a sensitivity loss of 17% which has only marginal impact for determining a delay screen and 100 s could therefore be taken as a representative ionosphere coherence time at 100 MHz.

More serious is the changing TEC difference observed by an interferometer that has a baseline length equal to half a wavelength when projected on the propagation direction of the TID. For a wave with amplitude 0.1 TECU and 90 km length propagating at ~ 150 m/s we get a maximum phase rate of 0.17 rad/s at 100 MHz. This leads for a 10 s integration interval to an amplitude attenuation by a factor $\text{sinc}(\delta\phi/2) = 0.88$, but much less degradation on projected baselines that are shorter or longer than half a wavelength. Increasing the integration time from 10 s to 14 s gives a degradation factor 0.78 but the noise reduces by a factor 0.85 giving a SNR that is reduced by a factor 0.92. However, estimating a phase rate from two successive 10 s samples allows a phase rate correction for every 1 s sample and integration over 20 s would increase the SNR by a factor 1.4.

This example shows the way for a tracking approach once a delay rate can be determined. Although the weakest source in the delay screen has $\text{SNR} > 3$, the strongest one is found in the central beam area and has $\text{SNR} > 9$, which allows establishing a rate of change over the integration interval for the reference direction by separating it into two half integration intervals. If this rate of change is a good first-order estimate also for the other weaker sources in the station beam longer integration becomes possible using a tracking procedure, up to the coherence time of 100 s at 100 MHz as defined above and proportionally shorter at lower frequencies. In this way the number of sources for which a delay and a delay rate can be determined is increased, but depends on the actual ionosphere conditions.

4.9 Summary of TEC screen modelling by self-calibration

The conclusions in previous sections and subsections based on the reported analyses can now be combined and summarized with reference to the subsections.

Summary of refraction and wavefront distortion aspects:

- The ionosphere is characterized by total electron content (TEC) along a ray path that causes excess delay, which is proportional to wavelength squared down to frequencies of ~ 20 MHz {4.1.1}. In normal conditions the TEC turbulence is limited and causes only refractive effects. Particularly during local sunrise diffractive effects can occur that cause amplitude effects, providing a clear indication that successful synthesis imaging is no longer possible.
- Observed interferometer phase is disturbed by differences in excess delay between the wavefronts towards two stations that form an interferometer and can consistently be described by a combination of medium scale Travelling Ionospheric Disturbances (TIDs) that “dissipate” into smaller scale Kolmogorov Turbulence fluctuations {4.3.3}.
- A simple (excess) delay screen model {4.2.1} consists of a curved thick homogeneous slab, a thin homogeneous wedge on top, and a thin delay screen at the bottom. All three elements describe phase shift proportional to the slanted excess delay and proportional to the secant of the zenith angle.
- Refraction by a wedge gives position shift proportional to the horizontal gradient of the excess pathlength in zenith direction and proportional to the secant squared of the zenith angle {4.1.4}.
- Spherical refraction by the curved slab {4.1.6} gives a position shift proportional to the excess pathlength in zenith direction, proportional to the secant squared of the zenith angle and proportional to the tangent of the zenith angle. A geometric derivation for an elevated homogeneous slab shows additional factors that become important at zenith angles larger than 45° . These terms are compared with the additional factors of a model using stratification and ray bending in the slab and differ mainly for zenith angles larger than 45° . A simple elevation independent additional factor gives however adequate description for elevation $< 45^\circ$. Also the differential refraction over the wide station beam ($< 10^\circ$) is then properly described for zenith angles of $45^\circ - 75^\circ$.
- Phased array stations do not need a pointing correction for refractions {4.1.3}. Only spherical refraction correction is needed in principle, but can be ignored in practice.

- A first-order description of the excess delay as function of frequency is no longer adequate below ~20 MHz and a second order approximation has been derived {4.1.1} valid for a slab with uniform distribution of the electron density.
- Typical medium scale TIDs with a wavelength of 90~200 km have a propagation speed of ~150 m/s {4.3.1} and the shortest ones induce the largest phase gradients for an interferometer when the projection of the baseline on the propagation direction equals half a wavelength.
- For a wave with amplitude 0.1 TECU and 90 km length we get at 100 MHz a maximum phase rate of 0.17 rad/s, which leads for a 10 s integration interval to an amplitude decrease by a factor $\text{sinc}(\delta\phi/2) \sim 0.88$ {4.8.4}.
- The degradation is less on baselines that are shorter or longer than half a wavelength when projected on the propagation direction, and integration times up to 100 s at 100 MHz can then be used as limited by propagation of Kolmogorov disturbances, and proportionally shorter at lower frequencies {4.8.4}.
- Refraction by a TID causes position shifts by the delay gradients but also higher order derivatives that cause image blur for an array larger than $1/12^{\text{th}}$ of a typical TID wavelength of ~90 km. Arrays larger than ~7 km need therefore for all stations further away than ~7 km from the centre of the array proper corrections for excess pathlength in each station beam as function of direction.
- Second order Lagrange interpolation using 5 reference sources that span an area with ~6° diameter gives TID interpolation errors over an area with 10% larger diameter that are smaller than ionosphere induced disturbances. This allows a station beam with 5° FWHM to be corrected down to 0.37, which covers $\frac{3}{4}$ of the solid angle {4.8.2}. Unfortunately the LOFAR LBA stations have larger beams leading to larger distances over which interpolation is required and correspondingly larger phase errors.
- This fixed maximum beam size is in contrast with beam matching to the characteristic coherence size according to the Kolmogorov model that would require at longer wavelength a maximum beam width that scales almost proportional to frequency to make the small-scale phase deviations per telescope independent of frequency.

Integrated source count data are the basis for estimating the number of available self-calibration sources per station beam and have been derived from published material {4.5}. Also we derived a relation that defines if sufficient flux is observed on long baselines that may partially resolve a potential source.

- Published differential source count data for 38 MHz to 1.4 GHz are analysed and combined to a single integrated source count covering 0.02 mJy - 20 Jy at 1.4 GHz. A power law description is used with seven intervals

- each with a fixed exponent, which gives an accuracy that is better than the statistical accuracy of the published segments {4.5.5}.
- The spectral index varies between 0.4-0.8 as function of 1.4 GHz flux and a value of 0.8 is adequate to derive for $S_{1.4} > 20$ mJy the source count for self-calibration sources at lower frequencies. For a 1.4 GHz flux below 0.02 mJy a spectral index of 0.7 is expected {4.5.5}.
 - The bi-modal size-flux relation at 1.4 GHz could be extended with published 324 MHz VLBI data showing that sources with $S_{324} > 80$ mJy and size $< 0.5''$ constitute a 25% subclass where the sources have an average spectral index $\alpha_{324}^{1.4} \sim 0.7$ and contain 1-3 narrower individual components {4.5.3}.

The number of available self-calibration sources per station beam that have a SNR > 3 within an ionosphere coherence time, given station aperture efficiency and available bandwidth, defines whether a TEC screen over each station beam can be derived that allows proper calibration of all weaker sources.

- Beam size and sensitivity of all LOFAR LBA stations are just adequate to find at 35 MHz on average 5 sources with SNR > 3 per interferometer in $\frac{3}{4}$ of the weighted beam area down to 0.37 of the peak sensitivity requiring 30 s integration time and $\sim 20\%$ relative bandwidth {4.6.2}.
- These 5 sources are enough to allow 2nd order Lagrange interpolation for each station beam that extends over ~ 22 km at ionosphere height {4.8.2}.
- The stations within a distance of 22 km from each other share a part of their spanned delay screens. Especially the screen area spanned by the core stations has therefore a large number of piercing points, which allows a much finer spatial sampling than is present in the screen over remote stations at larger distance from the core {4.7.6}.
- The European stations have double sensitivity but need the baselines to the LOFAR core to avoid resolving the sub class of sources smaller than $3''$ that constitutes $\sim 1/3^{\text{rd}}$ of the potential self-calibration sources and then have 5 sources per beam to span a curved delay screen {4.6.2}.
- However, the beam width of $\sim 9^\circ$ FWHM (at 35 MHz) is too large to give 3° sampling for a short TID wave. Quite fortunately, 6 additional beams with full sensitivity can be formed that surround the central one increasing the sensitivity at half power level and consequently the number of self-calibration sources {4.6.3}.
- The Dutch LBA stations have at 70 MHz a beam width of $\sim 9^\circ$ FWHM, since only the central area of the station aperture is used. An integration time of 20 s is adequate to provide 5 sources with 20% relative bandwidth. The remaining signal processing bandwidth allows two additional beams with full sensitivity or more beams with lower sensitivity that could still span the full band, but only sparsely. In this way additional self-calibrations

sources can be observed that not only extend the delay screen but also improve the sampling density{4.6.3} .

- The European stations have double sensitivity but at 70 MHz only a quarter of the beam solid angle while only $1/3^{\text{rd}}$ of the sources $< 1.5^{\circ}$ is not resolved on baselines to the LOFAR core stations. So, about 100 s integration time is needed to observe 5 sources over the beam that gives adequate sampling of a short TID. The processing power of the two additional beams could then be used to provide additional beams such that more sensitivity is obtained at half power level allowing a reduction of the integration time to give a better match to non-ideal ionosphere conditions {4.6.3}.
- At 140 MHz all stations have full effective aperture and need only 10 s and 20% relative bandwidth to observe ~ 36 sources in the central part of the beam where core and remote stations have an average sensitivity that is 0.78 of the peak value. The more sensitive European stations have a narrower beam but of all sources only a $1/4^{\text{th}}$ is not resolved on baselines to the LOFAR core, providing 9 sources that give adequate sampling of the delay screen {4.6.2}.
- The maximum number of sources per beam depends critically on the sparseness of the station as is demonstrated by the differences for LBA and HBA stations {4.6.2}.

Accuracy of peeled self-calibration phase parameters:

- The theoretical maximum number of parameters that can be solved per station per snapshot is determined by the number of independent baseline samples with this station and depends not only on the total number of stations but also on relative bandwidth and integration time {4.4}.
- The strongest self-calibration source in the beam does not suffer from noise induced phase unwrapping uncertainties and allows solving for the delay difference between stations including a potential residual phase term left after instrumental pass-band correction per station {4.7.1}.
- When all visibilities are corrected by self-calibration on the nominal position of the strongest source, not only the instrumental effects such as clock offsets are removed but also the differences between stations due to ionosphere excess delay. As a result a flat ionospheric excess TEC screen is defined over the array for this reference source direction.
- After subtraction of the nominal reference source structure from the corrected visibilities further “peeling” of the weaker sources is possible provided that they have $\text{SNR} > 3$ per baseline. Each “peeled” source with a known position relative to the strongest reference source is not only removed from the visibility data but provides a station based TEC value con-

- taining an arbitrary offset {4.7.4} but no contamination with direction independent station effects.
- The solutions found for reference source and first “peeled” source are contaminated by all weaker sources that have not yet been “peeled” away and require an iterative process and a bias correction. The published procedure [Tol, 2007] has been shown to be bias free if indeed all sources in a model simulation are solved leaving only the thermal noise {4.7.5}.
 - In practice there is a large number of weaker sources with $\text{SNR} < 3$ that cannot be “peeled”. We estimated the impact of these sources on the solution of the weakest source that could be “peeled”. We have shown that the thermal noise still dominates as long as less than 10 sources are “peeled” but we require that the baselines used in a decomposition are > 1 km while the relative bandwidth is $> 20\%$ {4.7.5}.

Thus far, it has been assumed that a TEC screen over the synthesis array could be constructed if TEC values could be determined in a number of directions of each station beam. We identified the various refraction effects by such a screen and analysed that such a screen could be constructed in principle from LOFAR data. Station based TEC values contain an arbitrary offset for each direction since each set of station parameters is derived independently from interferometer data that observe phase differences between stations.

An important result is that we identified a renormalization procedure {4.7.6} for these sets of station values that could provide a “smooth” screen spanned by station based TEC values defined for the zenith direction. Interpolation between the various positions and secant correction for the actual zenith angle allows a proper station based phase correction for every source direction and frequency.

When two station corrections are combined to correct an observed visibility any residual station offset will drop out. Potential bias in the corrections is limited {4.7.5} and interpolation errors are smaller than deviations induced by Kolmogorov turbulence {4.8.3}.

- The observed TEC at the reference positions in the screen needs appropriate correction for the inclination of the station rays {4.7.2} for use as TEC screen value.
- The renormalization procedure restores part of the TEC differences between the stations that were made zero by the initial visibility corrections based on the strongest reference source in the station beam {4.7.6}.
- Second order Lagrange interpolation between reference points in a delay screen needs an average reference point separation of at most $1/8^{\text{th}}$ of a TID wavelength to describe the assumed sine wave pattern sufficiently ac-

curate {4.8.2}. For a medium scale TID with a wavelength as short as 90 km at a height of 215 km we need a required average angular sampling of $< 3^\circ$.

- After tip-tilt correction for the large-scale TID induced phase gradients over interval range $< 3 r_0$, corresponding to $< 4.8^\circ$ at 100 MHz, the smaller scale residual Kolmogorov turbulence disturbances give a TEC deviation almost proportional to the distance from the nearest reference source of $\sim 0.008 \text{ TECU deg}^{-1}$ (rms) which is larger than the maximum interpolation error {4.8.3}.
- In good ionosphere conditions the interpolation error and the residual Kolmogorov turbulence deviations are even a factor 3 lower (and r_0 larger) giving at 70 MHz phase differences of 0.06 rad (rms) between piercing points that have 1 km separation. The phase gradient scales roughly to 0.12 rad/km at 35 MHz and to 0.03 rad/km at 140 MHz respectively {4.10}.

Integration times up to the ionosphere coherence time of order 100 s at 100 MHz (and proportionally shorter at lower frequencies) allow construction of a phase screen using a tracking approach and provide a proper averaged phase for the integration interval.

- If we want to reduce amplitude degradation of imaged objects we need to correct for the appropriate phase change over the interval.
- It is further suggested that the delay screen data averaged over a ~ 10 min interval is adequate to derive a residual refraction coefficient from the change in refraction over the field as function of zenith angle and as function of frequency over 20% relative bandwidth {4.2.3}.
- Also the contribution by the large-scale wedge could be determined {4.7.3}.

For a synthesized snapshot image we could extend self-calibration even further by

- Combining nominal and residual refraction coefficients the true thickness of the delay screen can be determined {4.7.2}.
- Including the large-scale and the differential delay with a model for the Earth magnetic field allows establishing Faraday and differential Faraday correction {4.1.2}.

4.10 Main Conclusions

The main conclusion of chapter 4 is that direction dependent self-calibration for wide field synthesis imaging by LOFAR is possible. The calibration approach involves the construction of a TEC screen over the station beams that describes large-scale refraction effects as well as disturbances by TIDs and Kolmogorov turbulence.

We estimated an ionosphere coherence time of 100 s at 100 MHz and proportionally shorter at lower frequencies, which limits the detection sensitivity for sources that need to span the TEC screen. We can indeed find at least 5 sources per beam for the LOFAR stations to span such a screen using 20% relative bandwidth and integration times between 10 - 100 s.

A renormalization process has been identified that allows combining station based solutions for the TEC in each source direction from observed interferometer phases. Exact self-calibration for these sources is possible, and for weaker sources in between the reference ones Kolmogorov turbulence induces a phase error per source per station. This turbulence error is proportional to distance from the nearest reference source, proportional to wavelength and dominates over second order interpolation uncertainty by the TEC screen.

Such a calibration procedure is in the first place required to do high quality imaging at high resolution over a wide FoV, but is also required to limit the errors on the nominal side lobes of the strongest sources. These error side lobes could otherwise determine the effective noise floor in a synthesis image, although the nominal side lobes are removed by subtraction, which is the subject of the next chapter.

The basic requirement for wide field self-calibration is that station beams are narrow enough to describe the shortest TID structures that dominate the ionosphere TEC screen over the beam with a simple 2nd order interpolation model. As example, we analysed 2nd order Lagrange interpolation that corrects exactly at the location of the strongest reference sources in the beam and provides sufficient accurate phase corrections for all other objects in the beam. Appropriate sampling of a TID over a station beam means at least 3 sources in the central part of the station beam and 2 more in the first annulus down to 0.37 of the peak sensitivity. The 5 sources need an average separation of $\sim 3^\circ$, or an inverse density of 8 deg² per source.

Not only the size of an antenna station is then a critical parameter but also the aperture efficiency needs to be sufficient to observe at least 5 sources with SNR > 3 per interferometer per station beam within an ionosphere coherence time, while the bandwidth is limited to $\sim 20\%$ of the observing frequency. When the separation between stations is smaller than the extend of the station beam at the effective height of the ionosphere TEC screen, stations share sampling points in the screen over the synthesis array and the required number of sources per beam can be relaxed.

According to table 4.4 we can indeed find 5 sources per station beam but this needs ~ 30 s integration at 35 MHz and ~ 20 s at 70 MHz, which allows to span a TEC screen that supports 2nd order interpolation. Unfortunately, the average separation between the reference sources is larger than 3° increasing the maximum distance between an interpolated and a reference position. In table 4.7 we summarize the latter value and calculate the associated rms phase error in an area with radius indicated by the separation relative to the nearest reference source due to residual Kolmogorov disturbances after correcting for large-scale effects using the formula derived in subsection 4.8.3

Table 4.7. Average residual phase errors over beam of remote station

Frequency	Station	Δt	BW	separation*	δTEC^{**}	$\delta\phi^{**}$
[MHz]		[s]	[MHz]		[TECU]	[rad]
140	HBA _c	10	42	0.4°	0.0024	0.14
70	LBA _s	20	14	2.4°	0.014	1.7
35	LBA	30	6	1.9°	0.011	2.8

* Radius of inverse source density in central beam area

** In good ionosphere condition a factor 3 lower

Important to realize is that the phase error at the edge of the area is a factor 1.35 larger.

At 140 MHz the average Kolmogorov error for interpolated sources is lower than the thermal noise induced phase error per interferometer for sources with SNR < 3 and shows that the screen interpolation approach eliminates ionosphere artefacts, just as if every source was independently self-calibrated.

Unfortunately, the LBA stations are in fact too sparse to provide adequate sensitivity to sample the TEC screen dense enough for accurate interpolation. However, the 5 observable source span a TEC screen over the station beam that allows 2nd order interpolation for a TID, but the maximum interpolation distance to a nearest reference source gives a Kolmogorov phase error that is only acceptable in good ionosphere conditions. Increasing the integration time to ~ 100 s could for good ionosphere condition improve the density of self-calibration sources in the beam of remote stations and reduce the phase error for the sources far from the references sources.

The 9.8° wide station beam at 70 MHz has at a TEC screen height of 215 km an extent of 37 km, which means that individual ionosphere sampling points per station beam can be shared with those of other stations that are closer to each other than 37 km. Especially for a station closer than 20 km from the LOFAR core the piercing

point density within its beam is increased significantly and a larger fraction of the beam could provide high quality visibilities on the baselines with this station.

The best observing strategy for high quality imaging is processing only those observations where good ionosphere conditions prevailed and forget about recorded worse ones.

5 Sensitivity Limitations by Artefacts in Aperture Synthesis

In this chapter we will discuss two types of artefacts, (i) the nominal side lobes inherent to Fourier imaging with incomplete sampling of the aperture plane, and (ii) the deviations from these nominal side lobes caused by phase and amplitude errors in the observed visibilities due to calibration and imaging approximations. Limiting the magnitude of these artefacts is a primary design driver for the configuration of a synthesis array and for the calibration and imaging procedures that together define its ultimate sensitivity. The calibration accuracy over the station beam defines the differences from the nominal side lobes and is a design driver for the minimum size of a phased array station as discussed in subsection 4.8.4. In this chapter we will estimate the value of the additional noise by calibration errors in an image as fraction of the thermal noise and discuss the processing requirements to reduce side lobe contributions. The important practical issue is the number of strongest sources to subtract from the visibility data to ensure that the side lobes of all weaker sources contribute less than the thermal noise. This number will drive the processing requirements for image forming as discussed in section 3.7.

Fourier imaging creates an image of the sky that is convolved with a point spread function (psf) given by the Fourier transform of the weight distribution of the observed visibilities. In practice we have finite and incomplete coverage of the visibility domain which results in a relatively strong side lobe pattern in the psf. The side lobes of strong sources therefore mask the weaker sources, requiring some deconvolution process to make sources of interest visible. Imperfect calibration and processing limitations cause baseline dependent phase errors that are even different for each source in the field. For complex gain errors we get deviations from the nominal point spread function (psf) and deconvolution with the nominal psf will leave a noise floor in the image that could well be larger than the thermal noise.

The results in this chapter will be derived by averaging over independent U, V -samples, which is appropriate for thermal noise contributions. As shown in previous chapters, the complex gain errors by self-calibration and imaging are station based and small errors cause deviations in the psf that will be addressed in section 5.3. Instead of a detailed U, V -distribution of an actual observation we use simplifying assumptions about the array configuration to obtain reasonable first order estimates for the effects of limited complex gain accuracy by self-calibration and disturbing ionosphere. Error lobes, that are a fraction of the nominal lobes, show the importance of good U, V -coverage that provides a low nominal psf side lobe pattern. Even better is a complete U, V -coverage over a limited area that can be made uniform by appropriate weighting after which appropriate tapering can provide nominal side lobes at a specified level.

There are different types of deconvolution methods and the most common ones in radio astronomy use iterative subtraction of nominal sources either from visibility data or, as a first approximation, from image data and require proper calibration and a proper source model. Practical implementations use an iteration process that finds the strongest sources, calibrates and removes them such that weaker source structures can be found in subsequent steps. An important difference between the two methods is that subtraction in the visibility domain allows perfect removal of object and associated artefacts using a proper complex gain for each source in each visibility. Subtraction in the image domain uses a psf that is constant over the image and this process cannot properly handle distortions by ionosphere, non-planarity, aliasing and other numerical and arithmetic errors that vary over the field. Actual image forming packages start with some initial calibration and form an image in which the strongest point sources are identified to form an initial source model. In an iterative process the source model is extended and the calibration parameters are improved. This model, using the improved calibration parameters after each step, is subtracted from the visibility data and reduces the impact of wide-field imaging artefacts. For LOFAR a different calibration procedure has been adopted using a global sky model (GSM) to identify the strongest sources in the station beam of an observation [Nijboer, 2006]. A multi-source self-calibration procedure determines the calibration parameters for at least 5 source directions and interpolated calibration parameters are used to correct visibilities for all other sources in the observation model as summarized in section 4.4.

The masking of weak sources by the side lobes of stronger sources in a field is called side lobe confusion and will be further discussed in section 5.1 where a first order estimate is given for the number of sources that should be subtracted. The actual level of the side lobes in a synthesis image is a crucial parameter that will be analysed in section 5.2 for snapshots with a random array, and will be extended to include effects of bandwidth and integration interval. Using a first order estimate for the side lobe level we will give estimates for the number of sources to be subtracted to reach the thermal noise level and conclude with a discussion on processing implications.

In section 5.3 we will show the relation between small complex gain errors per element in a phased array and the errors on the nominal side lobes in the beam of that phased array due to multi-direction self-calibration. The analysis will be extended to potentially large phase errors as could be induced by the ionosphere. Section 5.4 summarizes our results and the main conclusions.

5.1 Confusion aspects in a synthesis image

Confusion is an aspect related to sensitivity and resolution of an observing instrument and first encountered in radio astronomy when parts of the sky were imaged with a scanning telescope.

Classical *confusion* occurs when there is more than one source in the telescope beam. For a beam area Ω_b , the *confusion limit* S_c is the flux density at which this happens as one considers fainter and fainter sources. For an integral source count $N(>S)$, i.e. the number of sources per steradian brighter than flux density S , the number of sources in a telescope beam Ω_b is given by $\Omega_b N(>S)$. A survey is said to be *confusion-limited* if the expected minimum detectable flux density S_{\min} is lower than S_c , where S_c is given by $\Omega_b N(>S_c) \sim 1$. This definition stems from sky imaging with a single beam instrument and involves particular procedures that determine S_{\min} . The same definition could be used for imaging with a synthesis array where a large number of array beams is formed simultaneously. It must however be realized that the psf of an array has next to a main beam Ω_m an integral over all side lobes Ω_s that is more substantial than for a single dish antenna. Clearly, the confusion limit decreases with narrower main beam and lower side lobes. This aspect is an important design driver for a synthesis array where the total collecting area of all stations defines, together with the calibration and imaging procedures, the sensitivity S_{\min} , while resolution and side lobe level are determined by the distribution of the stations. An alternative design criterion that avoids confusion by limited resolution is a choice for $\Omega_m < (n N(>S_{\min}))^{-1}$ with $10 < n < 50$ [Taylor, 2004], [Bregman, 1999].

In an array of antennas the signals could be added and the squared modulus of the sum defines the real power pattern, which is a function of direction of the received signals. Such a single output array antenna has a narrow main beam that can be used to scan the sky and suffers from less classical confusion, although the side lobe confusion is much higher. Alternatively the individual signal products could be made available as in a correlation array, which allows forming beams by combining complex correlated powers. A Fourier transform could make a whole set of beams that provide together an instantaneous image of the sky. The baselines between the antennas define the positions of the correlation samples in the U, V -plane, while each point samples data that is convolved with the sampling function of the interferometer.

The antenna pattern of an array with N_{st} stations is evaluated as the sum of N_{st} phasors. For the direction where all signals of unity strength arrive in phase, the signal power is proportional to N_{st}^2 , but for directions where the phases have a random distribution the power is only proportional to N_{st} . As a result the psf of a narrow band snapshot image with N_{st} stations that are sparsely and randomly distributed over the antenna area has typical rms side lobe level N_{st}^{-1} . Only close to the main lobe the phase distribution could have some regular structure that allows higher and lower side lobes.

Also a U, V -plane that is randomly and sparsely filled with N_u samples has a psf with rms side lobe level $N_u^{-1/2}$. Although an array with N_{st} stations has a total number of $N_b = \frac{1}{2} N_{st}(N_{st} - 1)$ independent baselines that are used in sensitivity calculations, there are $2N_b$ sampled positions in the U, V -plane for a sparse random array and we get $N_u^{-1/2} \sim N_{st}^{-1}$. Including the N_{st} autocorrelations adds one more position at the centre of the U, V -plane that is often not used in practice. In practical sparse arrays, a higher density of samples is obtained near the centre of the U, V -plane, and this space taper reduces the level of the side lobes near the main lobe.

The correlation of the voltage beams of the two stations of an interferometer determines the integration function that samples the U, V -plane. An array with a number of stations samples the square of this number in the U, V -plane, which could lead to full sampling in principle of an aperture plane area with a sparse station distribution. This full sampling is an important requirement to obtain a clean synthesis image. In practice the aperture plane is not fully sampled and Fourier inversion gives a psf with side lobes. Appropriate tapering at the expense of reduced resolution can reduce the side lobe contribution introduced by the finite extent of the observed aperture. The side lobes due to missing visibilities cannot be reduced, resulting in the pick-up of residual signals of many sources over the sky that give a flux contribution in addition to the flux observed in the resolution beam. More important than this bias effect is the fluctuation level in this bias, this could dominate S_{min} when observations are averaged to reduce the thermal noise.

We will consider the actual LOFAR situation to get some practical figures for sensitivity and side lobe level and start with snapshot images that will be averaged to reduce the thermal noise as well as the side lobe level.

The side lobes of the strongest source in a snapshot image could mask sources that are well above the thermal noise in the snapshot image but do not exceed the side lobe level of the strongest source.

An LBA station beam has at 35 MHz a central part of $\sim 34 \text{ deg}^2$ which according to the source count given in table 4.2 contains on average 3 sources with a flux stronger than 0.78 Jy at 1.4 GHz, which can be converted to 15 Jy at 35 MHz. The annulus with a diameter of 1.2 FWHM around the central beam area has the same area but lower sensitivity and contains ~ 2 such sources. When the sensitivity is sufficient to detect these 5 sources with $SNR > 3$ per interferometer, a self-calibration solution will be possible that supports accurate subtraction of these 5 sources in the visibility domain as discussed in section 4.7. Even a phase screen can be derived that allows more, but weaker, sources to be subtracted with limited accuracy as discussed in section 4.8. An array with N_{st} stations has $N_b = \frac{1}{2} N_{st}(N_{st} - 1)$ independent baselines and the thermal noise in an image equals the thermal noise per interferometer but reduced by a factor $\sim 0.7 N_{st}$. For $N_{st} \sim 40$ we get a threshold of $(15 \text{ Jy}/28) \times (5/3) = 0.89 \text{ Jy}$ for sources with $SNR > 5$. Using table 4.2 we find ~ 225 sources in an LBA snapshot image that exceed the threshold and could be identified and subtracted in principle. $SNR > 5$ is chosen as appropriate

threshold to exceed the noise by the non-Gaussian side lobe distribution. Just below 0.89 Jy we are in the regime where the exponent in the power law of the integrated source count is about -1 and there are 450 sources with $\text{SNR} > 2.5$. If all ~ 225 sources with $\text{SNR} > 5$ are subtracted we still have ~ 225 sources with $5 > \text{SNR} > 2.5$. These sources have an average $\text{SNR} \sim 3.8$ and only a few of them can be identified from a single snapshot image. Even more serious, the nominal side lobes of these $N_s \sim 225$ sources constitute an additional noise floor of $3.8 N_s^{1/2} N_{\text{st}}^{-1} \sim 1.4$ times the thermal noise in a snapshot image. As discussed in subsection 4.7.5 all weaker sources double the side lobe variance giving a total rms noise level of $(1 + 2 \times 1.4^2)^{1/2} \sim 2.2$ times the thermal noise. This suggests that we can only identify in a snapshot image the sources that exceed 5 times the rms noise of a non-Gaussian side lobe distribution. Including the factor ~ 2.2 we can identify only ~ 102 sources in an LBA station beam as defined above that exceed ~ 11 times the thermal noise. This example shows in the first place that all sources below 5 times the thermal noise in a snapshot image limit seriously the number of sources that can be identified in a single snapshot image if the narrow band psf for a sparse random array is assumed. In the second place it is shown that the required number of sources to be subtracted to reach the noise floor in a snapshot image cannot be identified from a single snapshot image.

Averaging a number of snapshot images reduces the thermal noise and the level of the side lobe noise if the snapshots have an independent side lobe distribution. If the average side lobe level decreases faster than the square root of the number of sources between 2.5 and 5 times the thermal noise level, we could reach a situation that subtraction of additional sources is adequate to bring the effective noise floor close to the thermal one. These additional sources are all the ones stronger than 5 times the thermal noise in the set of averaged snapshot images. It might even be possible to subtract less than ~ 225 sources from each snapshot to get a final synthesis image where we are limited mainly by thermal noise,

This simplified analysis shows two important aspects:

- The side lobe level in a final synthesis image is an essential parameter to identify the number of sources that should be subtracted from each snapshot dataset.
- A step wise process is needed where first the few strongest sources are identified and subtracted from an image that has initially a higher noise floor before a next set of sources can be identified and subtracted.

Practical implementations of synthesis image-forming use both aspects and have demonstrated that the non-thermal noise is at about the level as the thermal noise in a long synthesis image that uses only a subset of all LOFAR stations.

5.2 Side lobe level in wide band snapshot synthesis imaging

In the previous section we concentrated on the rms side lobe level of the psf over the station beam and derived an additional rms noise level by the contributions of all sources in the station beam. We have seen in section 3.3 that the side lobes of sources outside the limited FoV of a small Fourier transform that covers only the station beam or a facet thereof still give contributions by the side lobes of the array psf. The sources outside the station main beam are however attenuated by the side lobe pattern of the phased array station. The LBA stations of LOFAR are random sparse arrays with a side lobe level of $\sim N_{\text{el}}^{-1}$ where N_{el} is the number of antennas. Since $N_{\text{el}} < 10^2$ the side lobe level is high although the average side lobe pattern of a synthesis array with stations that have different antenna distributions is lower. This high station side lobe pattern gives little suppression of sources outside the station beam and could lead to a high noise contribution in an image by the array psf of sources all over the sky. In practice, this effect is attenuated by the finite bandwidth and integration time of the correlated visibility samples.

The attenuation by finite bandwidth and integration time have been addressed in section 3.2 and expressions for the interferometer response of sources at large distance from the fringe tracking centre have been derived. In practice relative bandwidth and integration time are chosen such that only small degradations are encountered for sources in the station main beam. The effects on images are discussed in various chapters of [Taylor, 1999] showing source broadening and amplitude reduction that increase with distance from the field centre.

In our case we are interested in the rms side lobe contribution in an imaged continuum field not only by the sources in that field but by all sources in the sky. This rms contribution could be addressed in terms of side lobes in the psf of a Fourier transform, or even better, as the sum of source responses at each image pixel.

We can describe the effect of finite bandwidth as a sum of scaled narrow band snapshot images over a range of frequencies. Another valid description is by the Fourier transform of a sum of scaled U, V -distributions. In both domains we have a convolution in radial direction where the amount of convolution increases proportional to distance. In the visibility domain the scaling is from the centre of the domain, but in the image domain the psf scales from each source.

The effect of finite integration time leads to tangential averaging increasing with distance in both domains.

A detailed analysis is complicated, since we deal with quasi-convolutions in both domains instead of convolution with a fixed pattern in only one domain. We therefore pursue a simplified approach that combines aspects from image and visibility domain to obtain a first order estimate for the rms noise contribution in a Fourier image by sources within the imaged field but also by sources further out.

5.2.1 Array configuration

The psf of a synthesis image is determined by the actual distribution and weight of the U, V -samples.

In Earth rotation synthesis the U, V -samples follow contiguous tracks that form regular structures and as a consequence also the psf will show regular structures. These could in principle lead to full U, V -coverage, which in combination with appropriate weighting and tapering could produce a very low side lobe level.

In fact we are only interested in the rms value of the side lobes that determines the noise contribution due to the sources in the field. The mathematical basis will be given in subsection 5.2.5. In practice, a few high side lobes that could emerge from a regular structure in the station distribution could dominate this rms value. Instead of a detailed array model we use a simple model that shows the characteristic features of a randomized synthesis array such as LOFAR or SKA.

Our model array has a core area with radius L_c where about half of the stations are located, while the other stations are placed out to a distance L_{max} from the core. As a result about a quarter of all baselines is shorter than $2 L_c$, about half of all baselines are between L_c and L_{max} , while about a quarter has length between L_{max} and $2 L_{max}$.

An interesting result is that the two sub arrays with only short or only long baselines have the same sensitivity of half an array, but half the baselines of intermediate length provide a synthesis image with 0.7 times the point source sensitivity of the full array. Adding the three images together could at best give the full sensitivity, but the three psf main lobes have equal peak height but large differences in width. It means that the images need an appropriate weight to get a decent average psf pattern.

An array with N_{st} stations with diameter D samples $\sim N_{st}^2$ points in the U, V -plane that is convolved with the aperture sampling distribution of the station pairs with effective diameter D . As a result the radius of an instantaneous completely sampled U, V -plane could be $L_c \sim \frac{1}{2} N_{st} D$ for a properly configured sparse array of stations. The aperture area A_a is a factor N_{st} larger than the collecting area A_c of all stations, which shows that a large number of small stations gives not only better instantaneous U, V -coverage but provides also a larger FoV than a small number of larger stations.

An exponential distribution of N_{st} stations along an *East-West* line gives an exponential distribution of $\sim \frac{1}{2} N_{st}^2$ baselines. For a relative bandwidth $\Delta v/v$ each U, V -sample with baseline length B gets extended by $\Delta B = B \Delta v/v$ in radial direction. Starting with $B_{min} \sim 2 L_c$ a contiguous set of U, V -samples could be obtained up to $B_{max} = L_c (1 + \Delta v/v)^n$ with $n \sim \frac{1}{2} N_{st}^2$. A 12 h synthesis could then fill a complete

U, V -plane with wide tracks up to a radius B_{\max} . Instead of placing the stations on an East-West line they could be placed in annuli with exponential growing radii providing 2-D snapshot imaging capability and still fill the U, V -plane completely after 12 h.

We have shown the four basic principles used in the original array design of LOFAR that defined a consistent set of parameters for total number of stations, station size, core size and maximum baseline that supports the envisioned astronomical applications. In a later stage the total collecting area had to be reduced, while the site locations had been defined. The important full U, V -coverage for the core area could be approximated by reducing the size of the HBA stations and increasing their number. The actual number of remote stations has become too small to give full U, V -coverage over the long baseline range at the intended relative bandwidth. As a result the actual U, V -coverage shows a number of gaps that provide a sub array psf with relatively large side lobes. The actual array psf could be considered as the difference between the psf of a filled array and the psf of an array consisting of gaps, where the gap array needs a proper weight. Although the latter does not contain signal or noise it contributes significantly to the rms side lobe level of the actual array since the nominal array has an intrinsic low side lobe level assuming that appropriate taper is applied.

Instead of addressing the rms side lobe level from an U, V -distribution formed by tracks, we start from individual 2-D snapshots with a random distribution of U, V -samples and analyse how the narrow band psf pattern evolves in a multi-frequency snapshot image of limited bandwidth less than 1% and duration less than 10 min as discussed in earlier chapters. The practical importance is given by the forthcoming shallow surveys to be done with LOFAR using only a few of such snapshots spread over time and using up to 20% bandwidth per final image set that will provide spectral index information.

5.2.2 Quasi-convolution effects by bandwidth and time integration

A snapshot U, V -distribution at a single frequency provided by an array of stations is called sparse if the separation between the samples is larger than the size of the samples as observed with the interferometers. A second snapshot is independent of the first set when all samples of its U, V -distribution have separations from the samples in the first set that are larger than the aperture sampling width. It means that averaging of the two snapshot images gives an rms side lobe level that is reduced by 1.4. If part of the U, V -samples have less separation they are no longer independent and the reduction factor of the rms side lobe level of the averaged psf is less than 1.4.

A small rotation of the array and a small frequency change in the second snapshot produce a second set of U, V -samples close to the first set. The U, V -samples in the second set have separations from the samples in the first set that are proportional

to their distance from the origin. The sum of the two visibility data sets could be described by the visibility distribution of the first set but quasi-convolved by a two point pattern where the distance between the two points increases with radius from the origin. The second snapshot image has a slightly rotated psf that is scaled in radial direction. Adding the two images hardly effects the main lobe of the psf and its surrounding side lobes, but the furthest side lobes get extended and reduced in intensity. This can be described by a quasi-convolution of the psf, where the convolution kernel extends with distance from the main lobe. We apparently deal with quasi-convolution of the psf in the image domain and of the sample distribution in the visibility domain, which suggests that there must also be some taper relation.

The sparse sample distribution of a snapshot can be filled using Earth rotation, and for continuum observations by extending the bandwidth. Hence the side lobe level of a snapshot observation depends not only on the number of stations in a snapshot image but also on the relative bandwidth in relation to the relative resolution over the FoV, on the duration of the tracking interval and on the actual distribution and weight of the samples.

We know the rms side lobe level of a snapshot image of a sparse U, V -distribution and want to find the rms side lobe level of the average of a number of snapshot images. We discussed two extreme cases and we continue with a more formal explanation for an intermediate situation.

An important aspect is that the additional visibility samples due to bandwidth and rotation are adjacent to the samples of a narrow band instantaneous snapshot. These adjacent visibility samples can be considered independent for a FoV with radius $|\Delta I|$ when their separation ΔU satisfies $(\Delta I \cdot \Delta U) > 1$. In that case, all phasors that build up a side lobe outside a field, with radius $|\Delta I|$ around the main lobe, differ more than 2π in phase and build a side lobe pattern that is different from the reference pattern. U, V -samples with larger separations from the reference pattern give a random contribution to side lobes closer to the main beam. This means that within the FoV not all side lobes of patterns that stem from marginally different U, V -distributions do average with the square root of their total number, but that other forms of averaging play a role.

Ultimately, when independent U, V -samples overlap and fill the full aperture, complete and uniform filling could be obtained by weighting the samples and very low side lobes could be obtained by an appropriate taper function.

Our first step is analysing how the average side lobe level in a narrow band snapshot image decreases with increasing bandwidth and with increasing integration time and whether that decrease is sufficient to reach the thermal noise floor in a final synthesis image when only a limited set of sources is subtracted.

5.2.3 Frequency averaging

We start our analysis in the image domain with the psf of a sparse random array of N_{st} stations and connect it to the U, V -domain to get a consistent picture.

A multi-frequency snapshot image with relative bandwidth $\Delta v/v$ is the sum of a set of narrow band images, where the psf of each narrow band image is scaled proportional to wavelength. Our multi-frequency synthesis image averages the scaled psf versions and gives a main lobe of the psf at each nominal source position that has for small relative bandwidths a FWHM that corresponds to the FWHM of the main beam of the psf for the centre of the band. The side lobe pattern expands proportional to distance from the main beam and averaging over frequency for a distant position means averaging over different side lobes. This averaging could for a point source in the centre of the field be described as a convolution with a block function in radial direction. A single lobe with width δR convolved over distance $\Delta R < \delta R$ is extended and reduced in intensity by a multiplicative factor F given by

$$F = (1 + \Delta R / \delta R)^{-1} \quad (5.1)$$

A series of side lobes from different frequencies at angular distance R from the main lobe is convolved over a distance ΔR given by $\Delta R = R \Delta v/v$ and inserting this expression for ΔR in (5.1) gives

$$F_{v,R} = (1 + (R/\delta R) \Delta v/v)^{-1} \quad \text{for } R < R_{v,max} \ll 1 \quad (5.2)$$

This formula is valid as long as overlap with side lobes in radial direction is avoided and defines a distance from the main lobe over which (5.2) can be used given by

$$R_{v,max} = \delta R v/\Delta v \quad [\text{rad}] \quad (5.3)$$

At this distance we have a maximum attenuation factor for the side lobes given by $F_{v,max} \sim 0.5$.

The previous analysis in the image domain has an equivalent in the U, V -domain, where the snapshot samples at individual frequencies have an extent in radial direction defined by the finite relative bandwidth.

In a station configuration with about half of the stations in a central core, about half of all baselines are formed between remote and core stations. We take the average of these baselines as a characteristic distance B_c . U, V -samples at baselines longer than B_c get a larger extent, while samples at shorter baselines get a smaller extent for a given bandwidth. If we consider the extent of the U, V -samples as a form of convolution, we expect as a result some taper over the psf in the image domain for which we already have (5.2) for the central part.

The characteristic distance B_c in the array configuration defines a characteristic side lobe resolution $\delta R_c = 0.3 \lambda / B_c$. We want to connect the radial extent of samples in the U, V -domain to the radius $R_{v,max}$ in the image domain by choosing $R_{v,max}$ equal to the half power radius of the beam of a station with diameter D . Since $R_{1/2} = 0.6\lambda/D$, we define an extent of the sampling over distance D for a characteristic relative bandwidth $\Delta v_c/v$ by inserting the values for δR_c and for $R_{v,max}$ in (5.3) giving

$$\Delta v_c / v = 1/2 \ D / B_c \quad (5.4)$$

Inserting the values for δR_c and for $R_{v,max}$ in (5.3) defines a decay of the psf side lobes given by

$$F_{v,R} = (1 + R / R_c)^{-1} \quad \text{for } R < R_c \ll 1 \quad (5.4a)$$

With

$$R_c = R_{1/2} \Delta v_c / \Delta v \quad (5.4b)$$

For $R > R_c$ we get increased attenuation by averaging of side lobes that have some amplitude distribution. As argued in subsection 5.2.2 these side lobes are independent and the rms value of the resulting side lobes in the average over the characteristic frequency interval is in that case reduced by a multiplicative factor F'_v given by

$$F'_{v,R} = 0.5 (R_c / R)^{1/2} \quad \text{for } R > R_c \quad (5.5)$$

The U, V -distribution of the baselines between remote and core stations is the distribution of the remote stations convolved with the distribution of the core stations. The result is a set of clusters where each cluster has a radius L_c equal to the radius of the core.

For relative bandwidth $\Delta v_m > \Delta v_c$ we find according to (5.2b) a reduced R_c and a sampling extent $D_m = D \Delta v_m / \Delta v_c$. There is an actual limit since this extent should stay smaller than B_{sep} , some average separation between the baselines in a cluster to avoid overlapping sampling. Full filling of a cluster with diameter $2 L_c$ with $1/2 N_{st}$ cells (the number of stations in the core) with diameter B_{sep} we get

$$B_{sep} = 2.8 L_c N_{st}^{-1/2} \quad (5.6)$$

When D_m would exceed B_{sep} we no longer satisfy the requirement of independent U, V -samples and associated independent side lobes that average with $(R_c/R)^{1/2}$ and we get a different decay function for the side lobes. This defines a maximum bandwidth for validity of the simple decay function

$$\Delta v_m = \Delta v_c B_{sep} / D \quad (5.6a)$$

For Dutch LOFAR with $B_{\max} \sim 80$ km we find $B_c \sim 20$ km and for a station diameter $D \sim 80$ m of the LBA at 35 MHz we find a characteristic relative bandwidth given by $\Delta v_c / v \sim 2 \cdot 10^{-3}$. The minimum baseline separation between samples in a cluster for core radius $L_c \sim 1$ km and $N_{st} \sim 40$ is $B_{sep} \sim 0.64$ km, which defines $\Delta v_m / \Delta v_c \sim 8$ or a maximum relative bandwidth $\Delta v_m / v < 1.6\%$.

The actual attenuation of the side lobe level in a multi-frequency synthesis image as function of bandwidth is more complicated but these first order results are indicative for the psf side lobe attenuation that can be expected:

- The side lobes in a multi-frequency synthesis snapshot image decrease with increasing distance from the main lobe.
- A slowly decaying reduction to 0.5 w.r.t. a narrow band psf is reached at a distance that depends on resolution and relative bandwidth.
- More distant side lobes decrease with the square root of distance and relative bandwidth.
- There is a maximum relative bandwidth defined by the minimum separation between baselines

5.2.4 Time averaging

Combining snapshot images with different sky orientations involves correction for Earth rotation. As discussed in section 3.5 a synthesized snapshot needs first order corrections for continuous shift and rotation. A long synthesis involves combining synthesized snapshot images that are corrected for foreshortening and where l,m-coordinates of these snapshots are back projected to a coordinate system fixed to the sky before intensities are averaged.

This two-step approach involves small rotations that support according to subsections 3.5.4 and 3.6.4 at most a duration of ~ 10 min for the synthesized snapshots. Earth curvature could limit the duration of such snapshots even further as discussed in subsection 3.5.3.

The synthesized snapshot image could be considered as a sum of shorter snapshots images too. The small rotation during the synthesized snapshot results in larger tangential shifts at larger distance from the main beam of the psf that are the equivalent of the radial expansion by relative bandwidth. We therefore assume a comparable effect on the reduction of the near side lobes of the psf by (5.1) reaching ~ 0.5 at a distance $R_{t,\max}$ [rad] from the main lobe. The parallactic rotation varies with latitude of the array and with position of the field that is tracked by the array but we can use a worst case value just as in subsection 3.2.2. We express $R_{t,\max}$ as fraction of the half power beam radius $R_{1/2}$ defined in the previous subsection and for duration Δt [s] of the synthesized snapshot (that uses samples with much smaller integration time) we get

$$R_{t,\max} = 6878 R_{1/2} (D/B_{\max}) / \Delta t \quad [\text{rad}] \quad (5.7)$$

This analysis in the image domain has an equivalent in the U,V -domain, where tracks are formed. In the previous subsection we defined a characteristic relative change of frequency that defines independent side lobes beyond a characteristic distance from the main beam of the psf for snapshots at different frequencies. We define a characteristic time interval Δt_c using (5.6) for $R_{t,\max} = R_{1/2}$ with $B_{\max} = B_c$ and for the same LOFAR situation we get $\Delta t_c = 28$ s. The attenuation at $R_{1/2}$ is 0.5 and for distances $R > R_{t,\max}$ from the main lobe we get attenuation given by a multiplicative factor that describes the reduction in the rms side lobe level after averaging over a number of independent side lobes

$$F'_{t,R} = 0.5 (\Delta t / \Delta t_c)^{-1/2} (R/R_{1/2})^{-1/2} \quad \text{for } R > R_{1/2} \Delta t_c / \Delta t \quad (5.8)$$

We see again that the taper over the sparse random array psf decays with $R^{-1/2}$ starting from radius $R_{1/2} \Delta t_c / \Delta t$ in the image domain that is related to the tangential clustering of samples in the visibility domain. Here we find a maximum duration for which (5.8) holds

$$\Delta t_m = \Delta t_c B_{\text{sep}} / D \quad (5.8a)$$

giving $\Delta t = 4$ min for a large LBA synthesis snapshot that fills the baseline cluster of each remote station with independent samples, which allows (5.8) to be used. Further reduction of the side lobe level requires independent clusters that need a larger separation in time and will be discussed in subsection 5.2.7.

5.2.5 Combining frequency and time averaging

Thus far the analysis on frequency and time averaging has been done in the image domain and started with a convolution approach for individual side lobes of a narrow band instantaneous snapshot image. When the convolution extends over more than a single lobe we changed to an approach using averaging of side lobes with an independent amplitude distribution that reduces the rms of the resulting psf side lobes. We needed however recourse to the U,V -domain to define proper characteristic scales for time and relative frequency as basis for defining independent U,V -samples to get independent side lobes.

Averaging these independent side lobes results in some taper that decays with $(R/R_0)^{-1/2}$ in the image domain for radial as well as for tangential clustering of samples in the visibility domain. The decay starts from a radius R_0 that is smaller than $R_{1/2}$ the radius of the station beam at half maximum when integration time and relative bandwidth are larger than the derived characteristic values.

Multiplying the two taper functions of independent variables leads to a decay by $r^{-1} = (R/R_0)^{-1}$ which is an upper bound for the $|J_1(r)/r|$ taper that would result by true convolution of the snapshot U, V -distribution with a pillbox. In the U, V -domain we have for the multi-frequency imaging of the synthesized snapshot a quasi-convolution with a rectangular function in radial as well as in tangential direction. At the characteristic radius this width is equal to the range of the U, V -samples and the product of two convolution functions of independent variables is a rectangular pillbox. In a first order approximation, we replace the quasi-convolution by a convolution assuming a constant width of a square pillbox. A further assumption is that the Fourier transform of a square pillbox has the same upper bound as the $|J_1(r)/r|$ function for a circular pillbox. For a rectangular pillbox, the steeper of the two functions dominates the decay, where both decay with the square root. We evaluate the effect of decay with $r^{-1/2}$ and with r^{-1} and take the worst result as first order approximation

The reduction in rms level of the psf side lobes by summing narrow band snapshot images over a range of frequencies and time intervals is realized by a quasi-convolution of the sampling pattern in the U, V -domain.

- This taper gives limited reduction near the main lobe till radius R_1 , has a part decaying with $(R_1/R)^{1/2}$ and a part decaying with (R_2/R) from radius R_2 .
- Radii R_1 and R_2 depend on the square root of integration time and relative bandwidth.

These results will be valid for a synthesized multi-frequency snapshot image for a maximum relative bandwidth of 1.6% and a maximum integration time of 4 min for LOFAR with 80 m stations. Combining such wide band synthesized snapshot images will be discussed in subsection 5.2.7.

5.2.6 Effect of sources outside the main beam

The rms side lobe noise in a synthesis image contributed by all the sources outside the station main beam is given by integration over the sky over source flux weighted with average station side lobe level ϵ_{st} and a decaying array psf side lobe level ϵ_{ar} according to

$$\Delta S = \{ \sum_{sky} (\epsilon_{st} S \epsilon_{ar})^2 \}^{1/2} \quad (5.9)$$

Instead of a side lobe distribution characterized by a constant rms value $\epsilon_0 = N_{st}^{-1}$ over its extent we assume a decaying one with radius r from R_0 till R_{max} while S and ϵ_{st} are independent of r we get for a source density that is also independent of r

$$\Delta S = \epsilon_{st} < S^2 >^{1/2} \{ \sum_{sky} \epsilon_{ar}^2 \}^{1/2} \quad (5.9a)$$

with

$$\langle S^2 \rangle^{1/2} = S_{\text{rms}} = \{ \sum_{\text{sky}} S^2 \}^{1/2} \quad (5.9b)$$

where S_{rms} is the rms flux of all sources per square root steradian.

For $\varepsilon_{\text{ar}} = \varepsilon_0 (R_0/r)^{1/2}$ and integration over a circular area with radius r from R_0 till R_{max} we find

$$\Delta S = \varepsilon_{\text{st}} \varepsilon_0 S_{\text{rms}} (\pi R_0^2)^{1/2} (2 (R_{\text{max}} - R_0) / R_0)^{1/2} \quad (5.10)$$

For $\varepsilon_{\text{ar}} = \varepsilon_0 R_0 / r$ and integration over a circular area with radius r from R_0 till R_{max} we get

$$\Delta S = \varepsilon_{\text{st}} \varepsilon_0 S_{\text{rms}} (\pi R_0^2)^{1/2} (2 \ln(R_{\text{max}}/R_0))^{1/2} \quad (5.11)$$

We take for the contribution by sources outside the main beam $R_0 = R_{1/2}$ and for the LBA at 35 MHz we have $R_{1/2} \sim 4^\circ$ while beyond $R_{\text{max}} \sim 60^\circ$ the sensitivity of the stations is seriously degraded so contribution from there can be ignored. Interestingly, $(2 \ln(R_{\text{max}}/R_0)) \sim (2 (R_{\text{max}} - R_0) / R_0)^{1/2}$ for $2 < R_{\text{max}} / R_0 < 15$, showing that squaring the decay over an annulus introduces a square root in the last factor of the formula for the rms contribution of that annulus.

We need to evaluate how many sources in the sky outside the main beam are strong enough to be self-calibrated and how many are weaker but could contribute to the side lobe noise by their far side lobes.

We have shown in section 4.6 that snapshots with the LBA array can be self-calibrated at 35 MHz using $\sim 1/2$ min integration time but we need 20% relative bandwidth to provide sufficient sensitivity for detection of 3 sources in the central 34 deg^2 of the station beam with $\text{SNR} > 3$ per interferometer. This $1/2$ min happens to be about equal to the ionosphere coherence time as well as the characteristic time defined in subsection 5.2.4. According to section 5.1 this sensitivity corresponds to sources stronger than 15 Jy when observed in the station main beam. The station side lobe reduction requires that sources outside the main beam need to be a factor N_{el} stronger for an LBA station with N_{el} antennas to allow self-calibration and proper removal. We find a threshold of 720 Jy at 35 MHz that is exceeded by only a few sources, Cas A, Cyg A, Tau A and Vir A and a few other sources that happen to fall in a strong side lobe. These few sources can therefore be properly self-calibrated and subtracted, and we need to estimate the contribution by all weaker sources in the sky.

In subsection 5.1 the rms source flux in a sky area of interest was evaluated in two steps starting with $S_{\text{rms}} = S_{\text{bin}} N_{\text{bin}}^{1/2}$ where S_{bin} is the average flux of sources over a flux interval $(S_{\text{max}}, 1/2 S_{\text{max}})$ while N_{bin} equals the number of sources in that flux inter-

val. In the second step it was argued that given the source count all sources weaker than $\frac{1}{2} S_{\max}$ increase the rms flux for an integrated source count with index -1 only by a factor 1.4, while all sources stronger than S_{\max} are subtracted from the visibility data.

We evaluate S_{rms} by integrating over all flux below 720 Jy at 35 MHz according to

$$S_{\text{rms}} = \{ \sum_S S^2 N_0'(S) \}^{1/2} \quad (5.12)$$

where $N_0'(S)$ is the flux derivative of the integrated source count in the area πR_0^2 . Table 4.2 gives the integrated source count for various flux ranges at 1.4 GHz. The maximum flux of 720 Jy at 35 MHz corresponds at 1.4 GHz with 37 Jy and we integrate over 4 ranges starting at 20 mJy providing $S_{\text{rms}} = 48 \text{ Jy sr}^{-1/2}$. The contribution to S_{rms} by the intervals below 0.02 Jy is less than 0.6% and can be ignored. Over the used flux range we can according to table 4.2 assume a constant spectral index of 0.8 and convert S_{rms} back to a flux level of 912 Jy $\text{sr}^{-1/2}$ at 35 MHz.

Our snapshot has a total relative bandwidth $\sim 20\%$ and gives a thermal noise of 5 Jy per interferometer in about $\frac{1}{2}$ min, which can be scaled to the noise in a snapshot with characteristic duration of $\frac{1}{2}$ min and characteristic bandwidth 0.2%. Such a narrower band snapshot image would have a thermal noise of 1.8 Jy in an LBA snapshot image with $N_{\text{st}} \sim 40$ stations. The side lobe noise by all sources outside the main beam would then be given by (5.11) where R_0 equals the half power radius of $\sim 4^\circ$ at 35 MHz and $\varepsilon_0 \sim 0.5^2 N_{\text{st}}^{-1}$ since both frequency and time averaging reduced the side lobe level by 0.5 at that distance from the psf main lobe. With $\varepsilon_{\text{st}} \sim N_{\text{el}}^{-1}$ we get for $N_{\text{el}} \sim 48$ a value $\Delta S = 0.04 \text{ Jy}$, which is much smaller than the thermal noise of 1.8 Jy.

Averaging such snapshots to a synthesized multi-frequency snapshot of 4 min with 1.6% relative bandwidth reduces side lobe contribution and thermal noise both with a factor 8. Larger relative bandwidth and longer synthesis will reduce the side lobe level further but this decay could be slower than the decrease of the thermal noise, especially when large unfilled areas exist in the U, V -distribution.

We used a few simplifying assumption about the psf pattern to get a first order estimate for the contribution of sources outside the main beam of a station to the side lobe noise in a small image. We found that integrating over the rms source flux gives for a decay by $R^{-1/2}$ a result that is a factor 2.5 higher than for a side lobe decay by R^{-1} . The derivation assumed that all sources in the whole sky would be present in the correlated visibilities at their nominal strength. In practice the signals are attenuated by bandwidth and integration time decorrelation as discussed in section 3.2. Interestingly, fast facet imaging uses longer integration time and bandwidth per visibility sample that increase the attenuation by time and bandwidth decorrelation even for sources inside the station main beam.

We finally conclude:

- Side lobe noise by sources outside the main beam of an LBA station can be ignored.
- Such noise is even lower for the HBA with lower station side lobes and narrower main beam.

5.2.7 Combining snapshots in a synthesis image

For a synthesized snapshot with limited bandwidth and duration that are larger than the characteristic relative bandwidth of 0.2% and the characteristic integration time of $\frac{1}{2}$ min for LOFAR with large LBA, we expect side lobe reduction as derived in subsection 5.2.3 and 5.2.4 respectively. The relative bandwidth for narrow band continuum observations is typically limited to $\sim 1\%$, but a number of these could be combined to a wide band one of $\sim 20\%$ and a spectral index map. However, total integration time could span a rotation up to a full circle, which means that we cannot expect the tangential convolution analysis to be valid for this large range. This indicates that reduction of the side lobes with the square root of integration time has limitations. On the other hand, in an extreme case where more synthesized snapshots are combined in a long synthesis observation more than complete sampling of the U, V -plane could be obtained in principle. With an appropriate weighting scheme a uniform filling could be obtained over a contiguous area up to some maximum baseline. In such a case a point spread function with very low side lobes could be obtained in principle using an additional taper function. This will increase the thermal noise, but the non-thermal noise could almost be eliminated leading to a more sensitive image.

For a nearly filled U, V -plane we can consider the side lobe level as the difference between a low one of a fully filled and tapered distribution and a pattern formed by the psf of the gaps. The two patterns need a weight according to their filled area. It is therefore important that the gaps are randomly distributed since a large number of small areas gives a lower rms side lobe level than a few large ones. This reasoning shows that long gaps between U, V -tracks as a result of limited relative bandwidth could determine the side lobe level of a multi-frequency synthesis image.

When less than full U, V -coverage is possible, it is important that the total integration time is distributed such that the snapshots provide a U, V -distribution close to uniform but still random. In this case we can use the $N_u^{-1/2}$ formula. Relative bandwidth of $\sim 1.6\%$ fills the cluster of U, V -samples for each remote station in radial direction and in tangential direction after ~ 4 min tracking. By rotation we could fill a track at radius B_c by a number of sections with cluster diameter $2L_c$. In fact, we do not need continuous tracking, since after 4 min the cells in a cluster start to overlap instead of filling gaps. We assume only reduction by the square root of the increase in the

number of independent clusters and the attenuation factor will not be smaller than $(\frac{1}{2} \pi B_c / L_c)^{-1/2} \sim 0.18$ for the LBA case.

We summarize with results from previous sections:

- The reduction of the far side lobe pattern by bandwidth and total time of a synthesized multi-frequency snapshot psf with a sparse random array can in a first order approximation be described by a quasi-convolution of clustered samples in the U, V -domain as well as by a quasi-convolution of the side lobes in the image domain.
- The radial quasi-convolution by frequency averaging as well as the tangential quasi-convolution by time progression produce both a radially decaying taper over the side lobe pattern of the sparse random synthesis array.
- After a slow initial decay $(1+R/R_0)^{-1}$ further decay is bound by $0.5 (R/R_0)^{-1/2}$ for $R > R_0$.
- The radius R_0 at half value of the taper is defined by the characteristic resolution of the array and by relative bandwidth or by integration time. The characteristic resolution is determined by the baselines between core and remote stations and by wavelength.
- For two equal decay functions with R_0 equal to the radius $R_{1/2}$ of the station beam at half maximum, we find for the LOFAR with large LBA a characteristic relative band width $\Delta\nu_c / \nu \sim 0.2\%$ and a characteristic integration time $\Delta t_c \sim 0.5$ min that is accidentally well matched to ionosphere coherence time used for self-calibration of snapshot images.
- Longer duration and larger bandwidth have maxima expressed as multiples of their characteristic values given by B_{sep}/D for station diameter D . The average separation B_{sep} between the visibilities in the cluster area formed by the baselines from a remote station to the core stations is defined by core radius L_c and number of core stations. The maxima define the validity range of the derived decay functions, since overlap of sampling in the cluster is avoided.
- For the large LBA we find $B_{\text{sep}}/D \sim 8$ leading to a maximum relative bandwidth of 1.6% and a maximum duration of 4 min for which decay functions and their rms integration have been derived.
- The average side lobe level is mainly reduced by shrinkage of the area around the psf main lobe where the side lobes decay only slowly to 0.5^2 at radius $R_0 = R_{1/2} D/B_{\text{sep}}$.
- This is an important aspect in facet imaging, that relatively strong side lobes, as defined by the nominal psf of a narrow band snapshot dominate the field. In addition, the pick-up from neighbouring facets is reduced.
- Combining independent synthesized multi-frequency snapshots leads to equal reduction of thermal noise and side lobe level, when U, V -patterns do not overlap. The maximum reduction is given by $(\frac{1}{2} \pi B_c / L_c)^{-1/2} \sim 0.18$ for the large LBA case.

- Long contiguous tracks do not satisfy this independency requirement for the psf and build up systematic patterns. These patterns limit reducing the side lobe level while the thermal noise decreases.
- Ultimately, when the sampling over long time and wide frequency intervals fills the U, V -plane completely with sample distances smaller than the station diameter, a very low side lobe level could be obtained by appropriate weighting of all samples and by an appropriate taper function.

For incomplete U, V -coverage the disturbing side lobes of the strong sources can only be reduced by subtracting them accurately from the U, V, W -data before Fourier transformation of projected and corrected U, V -data. This drives the processing cost for subtraction and we will give estimates for the number of sources that have to be subtracted for different side lobe levels.

5.2.8 Minimum number of source subtractions

We want to estimate the number of source subtractions in an observation of 12 h having 1% bandwidth as a representative example for a narrow band continuum observation. In survey programs a number of such bands will be combined to a more sensitive wider band image spanning ~20% together with an image showing the spectral indices of all the sources. Such a 20% band is essential for LBA observing to allow proper self-calibration as discussed in chapter 4.

We identified three regimes for bandwidth and duration of a synthesis observation that govern the noise contribution by side lobes. We start with a psf for a narrow band snapshot image of a sparse random array with N_{st} stations that has a side lobe distribution characterized by an rms value $\varepsilon_0 = N_{st}^{-1}$ and we assumed this rms constant over the psf extent. For a multi-frequency synthesized snapshot image we derived a reduction of this rms value with distance from the main lobe. Equal reduction by time and bandwidth to $\sim 0.5^2$ is reached at a distance equal to the half power radius of the station beam by $\sim 0.2\%$ bandwidth and by $\sim \frac{1}{2}$ min integration time respectively.

Squared averaging of $\varepsilon_0 (1 + R/R_0)^{-1}$ over an area with radius R_0 gives an rms side lobe level $\sim 0.6 \varepsilon_0$. For the product of decay by frequency and time and we estimate an average rms reduction by 0.4.

Further extension into the second regime to 1% bandwidth and 2.5 min integration reduces the central area with slow decay of the side lobe level to 0.25 at a radius $R_0 = 0.2 R_{1/2}$, i.e. $1/5^{\text{th}}$ of the station beam radius. Continued further decay by 0.25 (R_0/R) gives according to (5.11) an rms contribution of $0.25 \varepsilon_0 2.5 (R_0/R_{1/2}) (\ln(R_{1/2}/R_0))^{1/2} \sim 0.16 \varepsilon_0$ over the rest of the area within a station beam. Comparing the rms noise contribution from the area within R_0 with the contribution by the rest of

the station beam using S_{rms} as in (5.11) shows that the rest of the station beam dominates the total side lobe noise contribution in the image.

Interestingly, we find the same result as taking a constant rms over the beam and decrease it using the increase of bandwidth and duration assuming an inverse square root dependence for both.

Longer summing over an interval such that U,V-samples within a cluster do not overlap will give further reduction by $(2.5/4)^{1/2} \sim 0.8$ for synthesized snapshots of maximum duration of 4 min.

Further time integration brings us in the third regime to 12 h and could give a maximum reduction by $(\frac{1}{2} \pi B_c / L_c)^{-1/2} \sim 0.18$ for the LBA case with 80 m stations.

Combining all factors for 1% relative bandwidth and 12 h gives a first order estimate of $\sim 6.5 \cdot 10^{-4}$ for the rms side lobe level over the station beam using $N_{\text{st}} \sim 40$ LBA stations of which 20 remote ones define B_c and the 20 core ones define L_c .

For smaller stations such as the LBA at 70 MHz and the HBA the characteristic time and relative bandwidth reduce with station diameter, approximately by 0.5. The factor B_{sep}/D compensates and keeps not only the maximum duration and maximum relative bandwidth the same, but also R_0 . An extra factor is the frequency scaling for R_0 .

The reduction by longer integration originates from further shrinking of the area with slowly decaying side lobes near the main lobe of the psf. Mainly since we assumed independent side lobe patterns further away that average with the square root of the number of time intervals. Further increase of the bandwidth could potentially lead to full U,V-coverage and very low side lobes by appropriate weighting and tapering. However, when gaps between tracks are present and build up large scale patterns, the rms side lobe level could be determined by this contribution.

We consider the derived side lobe estimate of $\varepsilon \sim 6.5 \cdot 10^{-4}$ as the lowest side lobe level that can be reached by averaging a number of well-spaced synthesized multi-frequency snapshot images of 4 min.

For continuous tracking during 12h we get U,V-tracks that can no longer be seen as a random distribution and could in practice give higher side lobes by gaps in the U,V-coverage or lower side lobes if full coverage is obtained and appropriately weighted.

To see the impact of a higher and a lower side lobe levels we present results for a range of rms levels. We calculate in table 5.1 the noise levels for three observing frequencies and estimate the minimum number of sources to be subtracted to reach the thermal noise level for a number of rms side lobe levels.

Table 5.1 Minimum number of subtractions to reach single pol noise floor in 12 h synthesis

Frequency	35 MHz		70 MHz		140 MHz	
Station type	LBA		LBA (reduced size)		HBA	
Central beam area	34.2 deg ²		54.8 deg ²		9.1 deg ²	
Subbands	2		4		7	
Effective bandwidth	0.35 MHz		0.7 MHz		1.2 MHz	
σ_{therm} $\sigma_{1.4}$	16.4	1.27 mJy	7.9	1.07 mJy	0.24	0.076 mJy
$N(>\sigma_{1.4}/0.78)$ $N(>\sigma_{1.4}/0.49)$	2274	1617	4115	2825	13846	7461
$\varepsilon = 2 \cdot 10^{-4} N_{\text{sub}} N'_{\text{sub}}$	1.3	0.8	15	7	273	187
$S_{1.4}$	980		300		3.53	
$\varepsilon = 5 \cdot 10^{-4} N_{\text{sub}} N'_{\text{sub}}$	44	24	170	79	464	346
$S_{1.4}$	100 mJy		58 mJy		1.8 mJy	
$\varepsilon = 10 \cdot 10^{-4} N_{\text{sub}} N'_{\text{sub}}$	146	92	387	268	1071	842
$S_{1.4}$	33 mJy		20 mJy		0.5 mJy	
$\varepsilon = 15 \cdot 10^{-4} N_{\text{sub}} N'_{\text{sub}}$	327	202	701	486	5519	3019
$S_{1.4}$	15 mJy		10 mJy		0.15 mJy	
$\varepsilon = 20 \cdot 10^{-4} N_{\text{sub}} N'_{\text{sub}}$	368	240				
$S_{1.4}$	13 mJy					

Explanation in text

We use table 4.2 to convert the thermal noise level σ_{therm} at each frequency to an equivalent 1.4 GHz source flux $\sigma_{1.4}$ and estimate the number of sources that exceed the rms noise floor. We use $\sigma_{1.4}/0.78$, to correct for the average gain of the centre part of an assumed Gaussian station beam profile. The annulus with a diameter of 1.2 FWHM has the same area but lower sensitivity and uses $\sigma_{1.4}/0.49$. These reference numbers can be compared with the numbers that have to be subtracted.

We use table 4.2 to derive for level S a formula for the rms flux $S_{\text{rms}}(<S)$ [Jy sr^{-1/2}] matched to the $(\pi R_0)^{1/2}$ factors in (5.11) by all sources that are weaker than S. If we tolerate an rms excess noise contribution of 0.4 σ_{therm} the image noise level will increase by a factor $(1 + 0.4^2)^{1/2} = 1.08$ or 8% above the thermal noise floor.

Using weights for central part and the annulus that depend on the flux level regime we derive the flux level $S_{1.4}$ (referred to peak sensitivity of the station beam) that would give for an assumed side lobe level such an observed rms noise over centre and annulus together.

We need subtraction of all sources $N_{\text{sub}} = N(>S_{1.4}/0.78)$ in centre part and $N'_{\text{sub}} = N(>S_{1.4}/0.49)$ in the annulus using table 4.2.

The following conclusions can be drawn from the table:

- An extremely important result is that the required number of source subtracts depends strongly on the side lobe level and on the total number of sources per station beam.
- A side lobe level of 10^{-3} for a long continuum synthesis observation indicates that the $\sim 8\%$ strongest sources above the thermal noise floor have to be subtracted leaving all weaker ones to raise the noise floor by only $\sim 8\%$.
- A factor 1.5 higher side lobe level requires 2.2, 1.8 or 4.5 times more source subtracts for the LBA at 35 MHz, 70 MHz and HBA at 140 MHz respectively
- A factor 2 lower side lobe level requires 3.5, 2.6 or 2.4 times less source subtracts respectively.
- Interpolating results for a side lobe level of $6.5 \cdot 10^{-4}$ estimated for the LBA configuration with 20 core and 20 remote stations shows 109, 358 or 1023 source subtracts respectively.

These effects depend strongly on the sensitivity and on the number of sources at the noise level. Observations that are more sensitive, as obtained with the HBA, have many more sources per beam. In addition, we have a different regime of the integrated source count formula. The consequence can be dramatic as illustrated in the last column of table 5.1, showing the critical importance of a low side lobe level. For an rms side lobe level of 10^{-3} we need to subtract all sources with $\text{SNR} > 6$, but a level of $1.5 \cdot 10^{-3}$ needs subtraction of all sources with $\text{SNR} > 2$. Although subtraction itself is possible, we cannot identify all required sources in practice and in that case we have to accept a degraded noise level for continuum observation in total intensity.

Increasing the bandwidth from 1% to 20% reduces the thermal noise by a factor 4.5, which makes all sources at the $\text{SNR} \sim 1$ level in a 1% bandwidth observation in principle firmly detectable in an observation with 20% bandwidth. We exceed in that case the 1.6% maximum bandwidth for side lobe reduction, and can only expect a smaller reduction due to sample overlap in the clusters, as discussed in subsection 5.2.3. To get a side lobe noise of only 40% in the wide band image all 1% bandwidth images have to subtract to lower levels.

When the U,V-tracks are broadened and start to overlap, we get parts of tracks with double weight with comparable impact on the psf as gaps between tracks. In this case a sensitivity loss of 8% will require more subtractions, and might even not be reached at all. Appropriate weighting can reduce this effect.

5.2.9 Processing implications

Subtraction in the visibility domain requires application of corrections derived from self-calibration, which needs 4 CMA (one Complex Multiply Add operation needs 6 floating point operations) per source per visibility as argued in subsection 3.7.1.6. This figure has been confirmed for the calibration package developed for LOFAR where station based processing overhead can be ignored in practice because of the large number of stations and the very large number of spectral channels per baseline that are corrected. In subsection 3.7.4 we estimated the various processing requirements in image forming and concluded that convolution processing dominates by far over Fourier transformation for continuum imaging in bands wider than 0.3%. Source subtraction equals about the minimum convolution processing when ~20 sources are subtracted. This latter figure is a practical minimum needed to subtract at least 4 sources outside the main beam and at least the 5 self-calibrators in a station beam. Each source subtraction requires ~24 flop leading to a total of at least 480 flop per complex visibility for a filled and properly weighted and tapered U,V-plane. Including Fourier transformation and dominating complex convolution we typically need 10^3 flop per complex visibility for the image forming.

At least ~100 more sources have to be subtracted for LBA observations at 35 MHz using 1% relative bandwidth requiring $2.5 \cdot 10^3$ flop. Even ~1000 more sources have to be subtracted for HBA observations requiring $2.5 \cdot 10^4$ flop per visibility. This latter number is comparable to estimates based on older imaging and self-calibration packages [Cornwell, 2004], [Yashar, 2009] although these authors used fields with much fewer sources.

Alternatively, subtraction could be performed in the image domain using a nominal psf with limited extent in a long synthesis image. This option needs positioning of the nominal psf outside grid points requiring at least a 3x3 interpolation for which we estimate the equivalent of 2 CMA per pixel of the psf of each source in the final synthesis image. This option could be attractive if the number of pixels in the limited psf extent is smaller than half the total number of visibilities in a synthesis image that need 4 CMA per source. Subsections 3.7.1.1 and 3.7.1.2 show that a long continuum image will have even more visibilities than image pixels making the approach feasible from processing perspective.

We would need a convolutional correction for amplitude variation by the station beams and for phase variation by the ionosphere to get the same psf for all point sources in the field. Since a convolution correction is performed for the *W*-term

anyway this is no large additional processing effort. The psf of a point source at the edge of a field will still be distorted by non-planarity effects left after 2nd order convolution correction. All residual numerical and arithmetic effects including aliasing cause deviations from the nominal psf. If we take these deviations for the sake of argument at 10% level, we could still use psf subtraction for sources of limited strength.

A more important saving on subtraction processing could be made if only the central part of the psf would need to be subtracted. This has an important effect on the area just around each source where we have little decay, but (5.10) shows that each annulus outside R_0 with a fixed ratio between outer and inner radius gives the same rms contribution. Subtracting only the central part will therefore not reduce the rms noise in an image significantly.

An even larger saving in processing could be reached by a very low side lobe level as demonstrated in table 5.1 for the LBA at 35 MHz. Such a very low level could be obtained when the U, V -plane is sampled fully, i.e. when the distances between the samples are smaller than the station diameter. A weighting scheme as discussed in subsection 5.2.7 could decrease the side lobe level for a filled U, V -plane when an appropriate taper function is applied and is therefore an effective method to reduce the processing requirements for source subtraction.

5.3 Side lobe noise after self-calibration and source subtraction

In the previous sections we analysed the equivalent noise floor in a snapshot image by the side lobes of all the sources in that image and concluded that only a limited subset of strongest sources have to be subtracted accurately to detect all weaker sources that exceed the thermal noise floor.

In this section we analyse the residual side lobe noise of all sources that have been subtracted using imperfect self-calibration.

In chapter 4 we have shown that self-calibration for the LOFAR synthesis array can solve for the complex gain of at least the 5 strongest sources per station beam per ionosphere coherence time. This does not only allow the subtraction of these 5 sources accurately, but also allows estimating complex gain differences for other directions within the station beam induced by the station beam patterns and by ionosphere disturbances such as TIDs. In section 4.10 we have estimated the magnitude of phase errors due to evolution of Kolmogorov turbulence with angular distance from directions of the reference sources. In section 3.6 we discussed the amplitude variation over the station beam and have shown that the nominal beam shape could provide a proper interpolation formula for the amplitude of the 5 observed complex gain parameters.

5.3.1 Errors in nominal side lobes by array element based complex gain errors

The effect of station based complex gain errors has been studied [Wijnholds, 2006] and a formula has been derived for beam forming with an array with N_{st} stations giving equal signals that have each a gain factor with uncorrelated real and imaginary parts of Gaussian distributed errors per station. When the real parts with unit magnitude have no error and the zero imaginary parts have small errors, the variance is halved and the formula can then be used for small Gaussian distributed phase errors with rms σ_ϕ . Separating complex noise in phase and amplitude noise is important in view of the differences in magnitudes related to their origin, ionosphere and deviations from a nominal station beam respectively. The resulting error side lobe pattern $\sigma_{psf}(\mathbf{l})$ as function of direction cosine vector \mathbf{l} for the normalized psf $P(\mathbf{l})$ of the array is given by

$$\sigma_{psf}(\mathbf{l}) = 1.4 \sigma_\phi N_{st}^{-1/2} (N_{st}^{-2} + P(\mathbf{l}))^{1/2} \quad (5.13)$$

- Equation (5.13) describes the errors in the power pattern of $P(\mathbf{l})$, of a phased array station, and can be used to derive tolerances on the accuracy of placements of element antennas in an LBA station and of tiles in a HBA station.

The derivation has been made for beam forming where a weighted sum of N_{st} equal station signals is used, while our snapshot images use a weighted sum of all correlations between all elements. The latter situation allows the use of an independent weight for each baseline, and in general does not include the auto correlations; while the former approach allows only a weight per station. We simplify by assuming complex weights with unit amplitude that allow full station based phase control over the beam pattern of the array. Omitting the auto correlations and their noise contribution makes the integral over the psf zero and drives part of the side lobes below zero. In that case we need to add an appropriate offset to get a psf with all positive side lobes before (5.13) can be used and two relevant regimes can be distinguished for the rms value of the error on the side lobes

$$\sigma_{psf}(\mathbf{l}) = 1.4 \sigma_\phi N_{st}^{-3/2} \quad \text{for } P < N_{st}^{-2} \quad (5.13a)$$

$$\sigma_{psf}(\mathbf{l}) = 1.4 \sigma_\phi N_{st}^{-1/2} P(\mathbf{l})^{1/2} \quad \text{for } P > N_{st}^{-2} \quad (5.13b)$$

Equation (5.13a) describes those areas where the psf has near zero values and (5.13b) describes the stronger parts that define the dominating response by sources at other locations in the snapshot Fourier image.

Stations in a synthesis array that follow Earth curvature give non co-planar baselines leading to source position dependent phase errors for 2-D Fourier imaging, which have been analysed in subsection 3.1.6. These deviations are corrected

when sources are subtracted from observed visibility data. However, for imaged sources these effects could be considered giving a position dependent psf, although a simple 2-D Fourier transform of the weights gives off-hand a constant nominal psf. The phase deviations do not have a Gaussian distribution and (5.13) gives only a first order approximation for the deviation between the position dependent psf and the nominal psf.

5.3.2 Noise contributions by error side lobes

Applying (5.13b) to a sparse random array where the side lobes of $P(I)$ have an rms amplitude N_{st}^{-1} give a relative error on these side lobes of $1.4 \sigma_\phi$, and indicates that $\sigma_\phi < 0.7$ defines the range for small phase errors. Larger phase errors give error lobes comparable to the nominal side lobes. The weakest source used for self-calibration gives, according to subsection 4.7.5, a value $\delta\phi \sim 0.33 N_{st}^{-1/2}$ due to thermal noise and insertion in (5.13b) gives an error side-lobe contribution

$$\sigma_{psf}(I) \sim 0.5 N_{st}^{-1} P(I)^{1/2} \quad (5.14)$$

- These error side lobes remain after subtraction of a point source using a nominal psf.

5.3.3 Noise contribution by self-calibration

The weakest self-calibration source has $SNR \sim 3$ per interferometer and has an $N_b^{1/2}$ higher SNR of $\sim 2 N_{st}$ in a snapshot image, which means that the rms error side lobe noise $\Delta S_{error}(I)$ can be expressed as fraction of the thermal rms noise ΔS_{therm} in a snapshot image leading to

$$\Delta S_{error}(I) / \Delta S_{therm} = 2 N_{st} \sigma_{psf}(I) \quad (5.15)$$

Inserting (5.14) for the weakest self-calibration source gives

$$\Delta S_{error}(I) / \Delta S_{therm} \sim P(I)^{1/2} \quad (5.16)$$

Interestingly, this equation resembles the shot noise formula for optical detection using $\Delta S_{therm} = 1$ and counting $P(I)$ in photons.

The few stronger self-calibration sources have inversely proportional lower phase errors but their actual contribution to the side lobe noise in a snapshot image is the same due to their proportionally larger flux and (5.16) is also valid for these sources. The total contribution by M strongest self-calibration sources is therefore a factor $M^{1/2}$ larger than (5.16) and implies for the side lobes of a narrow band sparse

random array with amplitude N_{st}^{-1} that M should be smaller than N_{st} to make error side lobe noise by self-calibration lower than the thermal noise in a snapshot image. This reasoning implies that the phase errors that create additional noise in a snapshot image are independent of the thermal noise in the data, which created noise in the self-calibration solutions in the first place. We consider our estimate therefore as an upper bound that can be compared to a lower bound provided by a Cramer Rao Bound (CRB) analysis.

In his section 6.2, “Fundamental imaging limits” [Wijnholds, 2010], such a CRB analysis is done and it has been shown that noise due to self-calibration is at least an order of magnitude lower than the thermal noise. According to subsection 4.7.5, we satisfy $M < 0.5 N_{st}$ in practice since $M \sim 5$ (+4 sources outside the station main beam) while $N_{st} \sim 40$ showing:

- Our upper bound analysis is sufficient to indicate that the proposed self-calibration approach subtracting the 5 strongest self-calibration sources gives an rms contribution smaller than $(M / N_{st})^{1/2} \sim 0.35$ times the thermal noise in a snapshot image, increasing that by at most 6%.
- A published lower bound analysis shows however that self-calibration noise can be ignored in practice.

5.3.4 Noise contribution by phase screen calibration

All sources weaker than the weakest self-calibration source use interpolated gain parameters that have at least the same phase errors as the weakest self-calibration source as shown in subsection 4.8.1. Subtraction of sources using interpolated phase errors leaves therefore a flux variance at every point in the snapshot image given by summing of the squared side lobe flux over all sources using (5.14). The 5 strongest sources that are weaker than the 5 self-calibration sources that defined the phase screen are in the flux bin with $3 > \text{SNR} > 1.5$ per interferometer and have an average flux of about $2.2 \times 0.7 N_{st}$ times the thermal noise in a snapshot image. After subtraction of these 5 sources that need interpolated correction from the phase screen we use (5.14) with $P(I) = N_{st}^{-1}$ for a sparse random array and get an error side lobe noise contribution expressed as fraction of the thermal noise in the snapshot image given by

$$\Delta S_{\text{error}} / \Delta S_{\text{therm}} = 1.54 N_{st}^{5/2} 0.5 N_{st}^{-1} N_{st}^{-1/2} \sim 1.7 N_{st}^{-1/2} \quad (5.17)$$

Taking $N_{st} \sim 40$ the residual error side lobes of the 5 strongest sources that have to use interpolated phase screen parameters give an rms noise contribution of 0.27 times the thermal noise in a narrow band snapshot image. All weaker sources that might be subtracted have proportionally lower noise contributions and leave together at most the same error contribution. This assumes an integrated source count with exponent -1 and every factor 2 lower bin then contributes half the variance of

the next higher bin. As discussed in subsection 4.7.5 this converges to a total variance that doubles and leads to a total rms contribution of 0.38 times the thermal noise.

We conclude:

- Residual side lobes of sources subtracted after phase screen self-calibration in a LOFAR snapshot image using 40 LBA stations are lower than 0.38 times the thermal noise in that narrow band image and increase the thermal noise by a factor of at most $(1 + 0.38^2)^{1/2} = 1.07$ or by 7%.
- This result can be generalized by stating that any self-calibration scheme that subtracts sources using complex gain corrections derived from stronger ones increases the effective noise floor by errors in the side lobes that are not subtracted.

The phase screen parameters for LOFAR are derived from wide band observations with ~20% relative bandwidth. This means that individual narrower band images have a much lower relative noise contribution than given by (5.17). Averaging over independent ionosphere coherence intervals reduces thermal noise and error noise at the same rate.

5.3.5 Noise contribution by Kolmogorov evolution in the phase screen

Apart from the noise in the phase screen, the interpolated phases have an rms error that increases due to Kolmogorov evolution with distance from the nearest reference position. Table 4.7 gives the rms of phase differences between centre and points in an area around a reference position. The rms of expected phase differences between centre and rim of this area are according to the discussion in subsection 4.8.3 a factor 1.35 larger. As a result, for the LBA situation, these errors could reach values well beyond 0.7 rad for which (5.13) is no longer valid.

In case of large station based rms phase errors we take as first order approximation a psf side lobe pattern with rms value N_{st}^{-1} while the point source itself is blurred having a peak intensity reduced by $\exp(-\sigma_\phi^2)$ for a Gaussian phase noise distribution with rms phase noise σ_ϕ . For a uniform distribution of station phase errors that span half a turn we have $\sigma_\phi \sim 0.91$ predicting a reduction by 0.44, while the actual reduction is ~ 0.5 . A uniform distribution spanning $3/4$ of a turn has $\sigma_\phi \sim 1.36$ predicting a reduction factor 0.16, while the actually resulting peak is ~ 0.19 . Apart from the reduced main lobe, additional lobes of the same strength develop as expected for a Rayleigh distribution of the side lobes, and for a full turn distribution of the phase errors there is no longer an identifiable peak at the nominal position.

In practice, the situation is less bleak since stations with separations smaller than their main beam extent (at the assumed TEC screen height) share piercing points

that also serve as reference positions. For stations in and near the core of the array the effective interpolation area is therefore reduced and, more importantly, the maximum Kolmogorov phase deviation could be less than 0.7 rad. This means in practice that only the longest baselines can have a strongly reduced contribution to the flux of imaged point sources, leaving a reduced and broadened point source response.

We evaluate for this case the error side lobe noise contribution of the 5 strongest sources in the phase screen that are weaker than the weakest self-calibration source as discussed in subsection 5.3.4. According to table 4.7 the Kolmogorov phase deviations from the reference positions could on average exceed 1.8 rad in typical ionosphere conditions only for remote LBA stations. In that case the maximum error side lobes are equal to the nominal side lobes of a sparse random array, i.e. equal to N_{st}^{-1} , but the signal is washed out. Since only the baselines from the furthest remote stations of the Dutch LBA give this large contribution, only $\frac{1}{4}$ of all baselines is affected. These 5 sources have an average flux of $1.54 N_{st}$ times the thermal noise in the snapshot image and their worst case side lobe noise contribution is given by

$$\Delta S_{side} / \Delta S_{therm} = \frac{1}{4} N_{st}^{-1} 1.54 N_{st} 5^{1/2} \sim 0.9 \quad (5.18)$$

All weaker sources together increase this contribution by a factor 1.4 following the same reasoning as in subsection 5.3.4. Equation (5.18) uses the narrow-band side lobe pattern and ignores the side lobe attenuation for a wide band snapshot image that is the average of a number of narrow band images. We have shown in subsection 5.2.3 that bandwidth integration can reduce the narrow-band side lobe level of a wide band synthesized snapshot by $8^{-1/2}$ at most for the LBA configuration. This factor reduces the rms side lobe noise to 0.45 times the thermal noise increasing the image noise by 10% at most. Averaging many snapshots over independent ionosphere coherence intervals reduces noise of side lobes that are induced by ionosphere phase noise at the same rate as the thermal noise and does not change their ratio.

We finally conclude:

- Visibility contributions of sources at larger distances from the reference positions in a phase screen get washed out by phase fluctuation on long baselines. This leads to sources with lower peak intensity and wider and lower psf main lobe as determined by the still contributing shorter baselines.
- The nominal side lobe pattern of these sources when observed with a sparse random array is changed into a pattern different from the nominal psf, but has the same rms magnitude that contributes to the side lobe noise.

- Only the furthest remote stations of the LBA that do not share reference positions in their TEC screen with other stations could suffer these effects in typical ionosphere conditions.
- Assuming that only $\frac{1}{4}$ of the LBA baselines have phase errors larger than 0.7 rad in most of their beam area, the scattered source power by the long baselines is 45% of the thermal noise, leading to an increase in the image noise by 10% when the remaining sources are subtracted.
- The HBA array has according to table 4.7 phase errors that are much smaller than 0.7 rad and a residual rms side lobe contribution can be ignored.

5.3.6 Noise contribution by image phase errors

The maximum tolerated phase deviation for an object in a 2-D Fourier image is a parameter of choice that also appears in the calculation of maximum integration time and maximum bandwidth. A maximum phase deviation of ~ 0.3 rad then results after averaging over the sawtooth like phase pattern in a maximum amplitude degradation of 1.7 % of the visibility on the longest baselines for sources at half power of the convolving beam, which is considered acceptable. The same maximum phase deviation is now also considered acceptable for the visibilities in a synthesis image and defines a maximum FoV.

The saw-tooth pattern over time of the phase error has a maximum peak-to-peak value of ~ 0.64 rad, on the baselines with stations that have the largest non-planarity, for sources at the edge of the FoV. This edge is defined at a distance from the field centre where the facet beam or the station beam has its half-power value. The saw-tooth has an rms value of ~ 0.09 rad and gives only small side lobe errors. The saw-tooth patterns are for intrinsic non-planarity station based and so is the derived rms phase error allowing the same type of analysis as used in previous subsections. Inserting this value in (5.13b) gives

$$\sigma_{\text{psf}}(I) = 0.13 N_{\text{st}}^{-1/2} P(I)^{1/2} \quad \text{for } P > N_{\text{st}}^{-2} \quad (5.19)$$

Assuming that at most $\frac{1}{4}$ of all baselines is affected, the rms over the field is reduced by a factor 0.5. For a synthesized snapshot image with limited duration we have $P(I) \sim N_{\text{st}}^{-1}$ and we get

$$\sigma_{\text{psf}} = 0.06 N_{\text{st}}^{-1} \quad (5.20)$$

For an array with 40 stations we find for a single synthesized snapshot $\sigma_{\text{psf}} \sim 0.0015$. In a 12 h synthesis using a number of independent synthesized snapshot images of ~ 10 min the rms error side lobe pattern due to imaging phase errors is a factor 0.12 reduced to $1.8 \cdot 10^{-4}$, which is lower than other contributions. Unfortunately-

ly, this value does not decrease when observations are repeated to reduce the thermal noise. For such observations, we need to reduce the imaging phase errors, for instance by reducing the maximum duration of the snapshot images.

Also the accuracy of the first-order rotation correction during a synthesized snapshot image improves in that case. But the procedure for combining of snapshot images has to use an accuracy that should depend on the ultimately required sensitivity when a number of such images is averaged.

5.3.7 Averaging of independent snapshot images

When more snapshots are averaged that have independent ionosphere induced phase errors per snapshot, error side lobes average down with the square root of the number of snapshots, just as the thermal noise. This means that error side lobe effects in a long synthesis can be ignored if they can be ignored per snapshot and if the snapshot duration equals the ionosphere coherence time after which a new, and independent screen of independent phase errors is present.

Averaging a number of independent snapshots reduces the side lobe level by the square root of that number and reduces the thermal noise level by the same factor. Independent means in this case that U, V -samples are not replicated and are randomly distributed. As a result no additional sources need to be subtracted and the noise in the averaged snapshot has the same fraction of side lobe noise as the individual snapshots. In practice regular U, V -structures could emerge that cause well defined side lobe structures that reduce not at the same rate as the thermal noise. In that case more sources should be subtracted, or the side lobe level should be reduced by other means, or the ensuing artefacts could be calculated and used in the deconvolution.

The nominal side lobe pattern is determined by the distribution of U, V -samples which means that the side lobe level no longer reduces when a long synthesis observation creates samples closer than a station diameter. The U, V -plane is in that case oversampled and weights can be applied such that the U, V -plane gets sampled with uniform weight density. Although the SNR is reduced only slightly, the uniform distribution allows using taper functions that give very low side lobes in the psf of the final synthesis image.

The various rms contributions can be considered independent and their variances need to be added to find a final rms value.

5.4 Summary and Conclusions

We investigated subtraction of sources including direction dependent calibration and imaging distortions to remove all their potential side lobes from a synthesis image. Assuming perfect calibration and Fourier imaging of all remaining sources we have a nominal psf due to incomplete sampling of the U,V -plane. The rms side lobe level of this nominal psf together with the integrated source count for a solid angle defined by a station beam determines the number of strongest sources that have to be subtracted to get the noise of the side lobes of all weaker sources close to the thermal noise. This number is crucial and if larger than 20 it dominates the processing for source image forming in continuum imaging.

In section 5.1 we made a first order estimate of ~ 200 sources stronger than 5 times the thermal noise in an LBA snapshot image at 35 MHz that after subtraction would leave the side lobe noise of all weaker sources and create a level of twice the thermal noise. The increased image noise raised the question whether we can identify all the sources that have to be subtracted to reach the thermal noise level. The critical parameter that determines the actual number of sources that need to be subtracted is the rms side lobe level in a set of averaged snapshots. We suffer not only from sources within the station beam but also from sources further out. The full sky provides a large rms source flux that is reduced by the side lobe level of the station beams and by frequency and bandwidth integration of the correlated visibility samples.

Instead of analysing the psf of a long observation from U,V -tracks that are broadened by bandwidth, we started our analysis in section 5.2 with narrow band snapshot imaging with a sparse random array for which simple side lobe formulae exist that can be used as basis for a first order approximation for a real array. Averaging narrow band snapshot images to a longer multi-frequency synthesized snapshot image reduces the rms side lobe level with the square root of the number of snapshots and frequency channels as long as the final U,V -sample distribution is sparse and random. The thermal noise decreases at the same rate keeping the side lobe noise contribution at the same fraction of the thermal noise. In practice regular U,V -structures can emerge that cause well defined side lobe structures that do not reduce with the square root of the number of snapshots, but could be taken care of in different ways.

We introduced a simplified model of the station array configuration where about half of the stations are located in a core with radius L_c and all other stations further out. The expo-shell configuration has a distribution of stations over annuli with exponentially growing average radii, which could in principle provide full U,V -coverage in less than 12 h using sufficient relative bandwidth. The U,V -distribution of a snapshot is given by a pattern that is the autocorrelation of the station pattern. About half of the baselines are formed between core and remote stations and have a pattern

described by the remote station configuration convolved with the pattern of the core stations. This subset of baselines alone provides 70% of the array sensitivity and is characterized by an average baseline length B_c , and dominates the side lobe contribution of all background sources that could ultimately define the detection limit of the array for continuum sources. The associated U, V -pattern has $\sim \frac{1}{2} N_{st}$ clusters with diameter $2L_c$ filled with $\sim \frac{1}{2} N_{st}$ station cells of diameter D . There is an average separation B_{sep} between the cells in the cluster given by $B_{sep} = 2.8 L_c N_{st}^{-1/2}$.

Such an array configuration with N_{st} stations has a sparse random distribution of U, V -samples that provides in a narrow band snapshot image a psf with an rms side lobe level $\epsilon_0 \sim N_{st}^{-1}$. The pattern is dominated by a distribution of side lobes that have an amplitude distribution with rms value N_{st}^{-1} that could be lower close near the main lobe of the psf. Extending the frequency range extends each U, V -sample to a radial segment with a length proportional to relative frequency range and to baseline length. Extending the observing time extends each radial segment of frequency samples in tangential direction providing a set of close samples in an almost rectangular area around each nominal U, V -position. The U, V -pattern of a synthesized multi-frequency snapshot can now be described as a quasi-convolution of the nominal distribution with a small rectangular distribution that scales in size with radius.

We identified 3 ranges in bandwidth and integration time extension of a snapshot that reduce the side lobe level.

- The small range creates additional samples within the aperture cell defined by the station aperture and tapers the psf side lobes outside a radius $R_{1/2}$ equal to the radius of the station beam at half maximum. This regime is defined by the characteristic duration and relative bandwidth determined by $\Delta\nu/\nu = D/B_c$. Decay outside $R_{1/2}$ is given by $0.25 R_{1/2}/R$.
- The medium range creates additional samples that fill the area around the cells in the cluster formed by baselines between a remote station and all the core stations. The maximum time and maximum relative bandwidth are a factor B_{sep}/D larger than the characteristic ones.
- The radius of the central area around the psf main lobe with an rms level of $\sim 0.4 N_{st}^{-1}$ is smaller and decay of the pattern at larger distance is given by $0.25 (R_{1/2}/R) (\Delta\nu/\Delta\nu)^{1/2} (\Delta t_c/\Delta t)^{1/2}$.
- The long range, especially by medium range tracks creates only a limited number of addition clusters that fill the U, V -plane by wide tracks, and reduces the radius where R^{-1} decay sets in.
- Any form of overlap does not reduce the side lobes.

For observations shorter than 12 h providing less than full U, V -coverage, it is important that the total observing time is distributed such that the snapshots provide N_u independent U, V -samples in a distribution close to uniform but still random. In

this case we can use the $N_u^{-1/2}$ formula for the rms of the psf side lobes. By appropriate selection of observing time intervals we could fill a wide track at radius B_c with a number clusters with diameter $2L_c$. In fact we do not need continuous tracking since after duration Δt_m , cells in a cluster start to overlap, while we need a contiguous ring of filled clusters. We need independent clusters along such a track, and the attenuation factor will not be smaller than $(\frac{1}{2} \pi B_c / L_c)^{-1/2} \sim 0.18$ for the LBA case.

Combining snapshots such that contiguous tracks are formed could violate the condition of averaging independent sample clusters at random positions and could lead to a higher rms level.

On the other hand with sufficient bandwidth and long tracking full U, V -sampling is potentially possible, and appropriate weighting could give uniform sampling density for samples closer together than the station diameter. An appropriate taper function could in that case provide a very low side lobe level, although sensitivity and resolution will be reduced. The main result is that in such a case only tens of sources have to be subtracted instead of hundreds to reach the thermal noise in an LBA observation.

In practice, incomplete sampling of the U, V -plane is often not by a large number of randomly distributed small areas but could be formed by large track like structures. The impact on the psf can be estimated by considering the psf of the array as the difference between a nominal one with potentially low side lobes and one by overlapping structures or by gaps which create high side lobes. Appropriate weighting can cure the effect of overlapping structures, but gaps have to be avoided by an appropriate array configuration in combination with sufficient tracking time and bandwidth.

We reached an important result for facet imaging that uses small fields where the psf taper reduces the rms in the area around the main lobe to $\sim 0.4 N_{st}^{-1}$. The larger effective bandwidth and longer sample integration times per facet image reduce the radius where decay starts, strongly reducing contribution by sources of other facets within the station beam. Most sources need to be subtracted from visibilities in each facet but fast facetting has reduced the number of visibilities per facet, although the total number of visibilities of all facets together is not changed. As a result the total number of source subtractions in a station beam is reduced.

For Dutch LOFAR we have $L \sim 1$ km and $B_c \sim 20$ km which give for LBA stations with $D \sim 80$ m:

- $\Delta t_c \sim \frac{1}{2}$ min and $\Delta v/v \sim 0.2\%$ and a maximum relative bandwidth $\Delta v_m/v \sim 1.6\%$ and a maximum tracking time $\Delta t_m \sim 4$ min to fill the cluster of baselines per remote station.

Interestingly the characteristic time is about equal to the ionosphere coherence time and the maximum tracking time is comparable to the period for which a fixed polarization rotation correction can be used.

We derived a first order psf side lobe level of $6.5 \cdot 10^{-4}$ for a sparse random array with 40 stations with LBA configuration that observes for 12 h with 1% relative bandwidth typical for bands in a continuum observing that allow spectral index estimation. We estimated the number of source subtractions to allow an increase of 8% in the thermal noise floor of the synthesis image for a range of rms side lobe values in table 5.1. Interpolating our side lobe estimate for the LBA configuration shows:

- Subtraction of 109 sources for observing at 35 MHz and 358 sources at 70 MHz is sufficient, if we assume the same rms value for the smaller stations, respectively.
- According to table 5.1, these numbers equal 3 - 5% of all sources stronger than the thermal noise level in the final synthesis image.
- An LBA array that has a psf with a factor 2 higher rms side lobe level should subtract ~ 4 times more sources.
- Actually, we need 20% relative bandwidth, i.e. combining all 1% bands, to obtain proper self-calibration using the 5 strongest sources in the beam.

The more sensitive HBA array should in that case subtract 10 times more sources above a level that is two times the thermal noise, which cannot be realized and a higher final noise level has to be accepted.

In section 5.3 we investigated the effects of residual phase and amplitude errors after self-calibration and subtraction of the strongest sources using interpolated phases for the ionosphere and interpolated amplitudes for the station beams.

It has been shown that station based Gaussian distributed phase errors smaller than 0.7 rad give errors in the psf of a snapshot image with rms level equal to the product of the square root of the psf and the rms noise in the average of all station phases. This result has a number of important consequences:

- Tolerances can be set on the placements of elements in an array, such as a station, to limit deviations from a nominal psf.
- Noise by residual side lobes of subtracted self-calibration sources can be ignored for LOFAR.
- Noise by residual side lobes of subtracted sources that are calibrated using a phase screen spanned by 5 self-calibration sources in a station beam gives a side lobe noise contribution of at most 38% of the thermal noise and could increase the imagel noise by less than 7%.
- These statements can be generalized to other direction dependent self-calibration schemes that correct for large scale effects and leave only thermal noise induced effects.

- Averaging of snapshots with independent phase noise, as induced for instance by the ionosphere, reduces the error side lobe noise in a longer observation level with the square root of the number of ionosphere coherence intervals. Also the thermal noise is reduced that way and keeps the side lobe noise at the same relative level as in a snapshot with the duration of an ionosphere coherence interval.

Kolmogorov evolution causes station based phase errors that increase with distance from the reference points that span the phase screen and could reach values well beyond 0.7 rad leading to increased blur and reduced peak intensity as well as to additional side lobes.

- In practice only $\frac{1}{4}$ of the LBA baselines of the Dutch LOFAR suffer from such phase errors in typical ionosphere conditions and wash out signals on the long baselines.
- This will reduce point sources at increasing distances from the reference positions in the TEC screen to at most 70% of their peak flux.
- The flux that is scattered to side lobes gives a non-thermal rms contribution of order 45% of the thermal noise and could increase the image noise by 10%.

The main conclusions of the chapter are:

- The psf of a randomized synthesis array with about half the stations in a central core has three regimes for bandwidth and integration time that define the rms side lobe level.
- All sources outside the station main beam contribute an rms side lobe noise less than a few per cent of the thermal noise, after self-calibration and subtraction of the strongest sources in the sky, including Cas A, Cyg A, Tau A, Vir A and a few 3C sources using known source models.
- The 5 strongest sources in the station main beam can be self-calibrated and subtracted leaving an additional noise contribution that can be ignored as well.
- The self-cal solutions allow interpolated calibration for all other sources in the main beam that give errors in the nominal side lobes that contribute less than 7% to the image noise.
- Even good ionosphere conditions leave at 35 MHz large phase errors on the longest baselines for sources far from the reference sources. As a result a considerable fraction of the station beam observes these sources with limited resolution and flux. However, the scattered flux contributes a side lobe noise contribution of order 45% of the thermal noise.

- The number of sources that have to be subtracted using a nominal psf by calibration depends on the rms side lobe level, and on the requirement that the non-thermal rms contributions by all weaker sources is less than 40% of the thermal noise, contributing less than 8% to the image noise.
- Interpolated calibration and sources subtraction together contribute ~14% to the system noise in a synthesis image. The scattered noise in LBA images brings the total to ~23%.
- We estimated subtraction of ~100 sources for LBA observations of 12 h and 1% bandwidth at 35 MHz and much more for higher frequencies.
- HBA observations at 140 MHz could need even need more than 1000 source subtractions.
- Full U,V-coverage and appropriate weighting and tapering allows subtraction of less than 20 sources to reach the thermal noise within 10% and reduces processing for image forming to a minimum where subtraction and minimized convolution share the load.
- The additional side lobe level in a synthesized snapshot image, due to phase errors by non-planarity and first order field rotation corrections can be ignored in a 12 h synthesis image with more than 30 stations.
 - The error pattern is however systematic and when many images have to be added together to reduce the thermal noise, final image noise could become dominated by side lobe noise.
 - This systematic effect can be reduced by decreasing the FoV or by reducing the duration of such a synthesized snapshot.

The results in this chapter are derived based on reasoning from first principles and allow generic analysis of a synthesis array such as LOFAR using the simplest assumptions about its configuration. An essential non-trivial element in the analysis is the formula for noise in side lobes of an array snapshot psf induced by station based noise [Wijnholds, 2006], which allowed to address the effects of ionosphere induced phase errors that cannot be self-calibrated.

Our results form a well-documented reference for results from simulation and observation that contain effects that cannot be traced down easily, and define minimum processing requirements for thermal noise limited imaging at low frequencies.

A recent result [Yattawatta, 2012b] indicates that a 6 h observation with LOFAR using a total bandwidth of ~42 MHz reaches an image noise that is a factor 1.4 larger than the expected thermal noise. This shows that other noise contributions are about equal to the thermal noise. This relatively high value is according to our analysis to be expected for observations with more than 1.6% relative bandwidth. On the other hand, only ~500 sources are effectively subtracted, leaving room for additional subtractions.

6 Conclusions and Recommendations

The primary goal of the dissertation is a detailed analysis of the scaling laws that determine processing resources in an aperture synthesis array as function of station size and number of stations. The results and conclusions of previous chapters have been summarized in a separate section at the end of each chapter and will not be repeated here. Instead, they will be combined to more generalized statements, but lack the rigour of the original formulation.

In the low frequency range 10 - 240 MHz where LOFAR operates, the ionosphere induces large phase variations over angular separations of $\sim 12^\circ$ comparable to the width of station beams. Such a large field-of-view (FoV) requires not only multi-direction self-calibration on time scales defined by the ionosphere coherence time, but foremost sufficient calibration sources need to be observable per station beam with sufficient signal to noise ratio. Although LOFAR has an instantaneous bandwidth of 100 MHz, practical calibration is limited to a relative bandwidth of $\sim 20\%$ and imaging to even narrower bandwidth to avoid artefacts. Given integration time and bandwidth the sensitivity for detecting sufficient sources is determined by the density of antenna elements in a station. Although an important aspect in system design, it has been addressed only globally in chapter 2. LOFAR has been designed with these constraints in mind, and a detailed analysis on calibratability has been given in chapter 4.

LOFAR uses phased array stations that suffer from foreshortening, which introduces a varying amplitude variation over the FoV. The varying amplitude of sources and the varying ionosphere induced phase deviations cannot be handled by current legacy calibration and imaging packages that also show unfavourable scaling of processing requirements with FoV and resolution. These aspects have been analysed in chapter 3, showing that imaging can be optimized such that processing for image forming is proportional to the solid angle of the FoV of a station measured in resolution of the array.

We have shown that a fractional bandwidth of about 1% provides sufficient sensitivity with the HBA array to realize full direction dependent self-calibration over regions spanning less than 4° on the sky. The array of less sensitive LBA stations needs about 20% fractional bandwidth with a station beam that is about twice as wide, which allows full direction dependent self-calibration only in good ionosphere conditions. Such fractional bandwidths allow distribution of the total collecting area over a limited set of stations such that full sampling of the visibility distribution over the array aperture can be obtained using Earth rotation synthesis for periods of about 6 hours and longer. Multi-frequency synthesis effectively adds the set of narrow band point spread functions (psf) for each source to a wide band psf with much lower side lobes. Only when all sources have the same spectral index they get the same

wide band psf. In practice the visibility distribution is incomplete and side lobes could introduce additional side lobe noise that increases the thermal noise by sky and receivers. By subtraction of the strongest sources the side lobe noise by all weaker sources can be reduced at the expense of additional processing power for forming of continuum images, and has been discussed in chapter 5.

The LOFAR concept design presented in 1999 argued that this would not only be possible in principle, but could be materialized after 2003 when the cost of signal and data processing as well as the cost of wide band data transport would reach a level making such an endeavour feasible in practice.

Conventional imaging packages that calibrate and transform correlated visibility data into images have been developed for small fields at higher frequencies and lack efficient algorithms that can handle data volumes that are 10,000 times larger than conventional ones that are handled by a single laptop. Thus far, focus at ASTRON has mainly been on self-calibration, which needs to handle the disturbances by the ionosphere. These issues have been discussed in the calibration chapter where the aspects at the various scales in the ionosphere are summarized and put together in a consistent framework. However, development of high performance imaging packages has been left to the international community with its focus on higher frequency applications with much smaller FoV. In the imaging chapter the limitations of approximate 2-D Fourier imaging for the very large fields of LOFAR have been analysed and processing efficient algorithms have been proposed that scale proportional to FoV expressed in resolution elements. In the chapter on imaging artefacts we addressed the issue of sensitivity limitation by side lobes inherent to incomplete U,V-coverage and by the errors in these side lobes due to limited calibration accuracy. We started with the poor coverage by narrow band instantaneous images with a 2-D array and derived first order estimates for the side lobe level when bandwidth and duration of the synthesized snapshot images are extended.

The main conclusions of these three chapters can be summarized as follows:

- Processing for optimized continuum image forming is dominated by source subtraction if more than 20 sources have to be subtracted.
- When full U,V-coverage is available in a long wide band synthesis observation the side lobe level could become sufficiently low such that that order 20 sources need to be subtracted to reach the thermal noise level within 8%.
- Incomplete U,V-coverage has been analysed for an array with ~40 stations and LOFAR like distribution, where synthesized snapshots with 1% relative bandwidth and 6 min duration are combined for 12 h. First order analysis shows that fewer than ~50 sources need to be subtracted for observations at 35 MHz increasing to ~100 and ~400 sources at 70 MHz and 140 MHz

respectively, which dramatically drives the processing requirements to limit the noise increase to 8%.

- Wide field self-calibration using interpolated parameters from at least 5 reference sources per station beam smaller than 4° leaves residual phase errors that according to an upper bound analysis will increase the thermal noise in an image by less than 7%.
- LBA observing allows high quality high resolution imaging only in areas around the reference calibration sources. In the remaining area sources get blurred by reduced contributions of the baselines provided by remote stations, but the scattered source power increases the noise floor with 3% if the 'blurred' sources are properly subtracted. Stations with shorter baselines for which the beams overlap, provide additional areas with good calibration for these baselines. Good ionosphere conditions increase the size of the good areas.
- Foreshortening of the phased array station beam as well as its polarization characteristics are well behaved, and pose no foreseeable problems.

Higher side lobe levels require progressively more source subtractions consistent with actual numbers larger than 1000 at the most sensitive 140 MHz system, which drives the processing requirements by a factor 50 compared to an array that would provide a better U,V-distribution. The actual LOFAR array configuration suffers from gaps in the U,V-plane that increase the side lobe level to the extent where rms side lobe noise becomes comparable to the thermal noise in the longest observations, which is the most important message that this precursor instrument can give to the designers of the SKA.

The most important issue for the system design of the SKA is the distribution of the total planned collecting area over a number of stations such that the thermal noise can be reached with processing resources that use a reasonable fraction of the total system cost.

In following sections we summarize scaling laws for efficient processing of wide field images and limitations set by self-calibration that are key in the system design of future synthesis arrays using phased array stations operating at low frequencies.

6.1 Scaling laws in Fourier imaging

The most important final conclusion as presented in the introduction of this chapter is that for the smallest convolution kernel of 7^2 the processing for image forming becomes dominated by source subtraction if more than ~ 20 sources need to be subtracted. In this section we combine results of chapter 3 and chapter 5 how distribution of a total collecting area over a number of station affects the total processing for image forming.

The basis for efficient Fourier imaging is the use of the Fast Fourier Transform (FFT) that needs a regridding of observed visibility samples on a rectangular grid using a convolution process. A planar array could simply use a 2-D FFT to provide an image that accurately describes a hemisphere, but Earth curvature causes deformations that can be partially corrected by a complex convolution. Earth rotation has the effect that in a projected hemisphere objects move with different rates that limit the duration of a synthesis image. For a small field proper imaging can be realized by first order corrections such as rotation and projection of the baselines on a reference plane and by a second order correction with a complex convolution for the non-coplanar baselines. Unfortunately large projection angles cause third order effects for these non-coplanar baselines that limit the duration of a synthesized snapshot image that uses the plane of the array as the reference plane for 2-D Fourier imaging. This is the main reason for most imaging packages to use a reference plane perpendicular to the line of sight to a field centre defined on the sky, and deal only with the much larger non-planarity of an Earth bound 2-D array that tracks the sky field. The advantage of this approach is that the third order term disappears and only a single FFT is needed irrespective of the duration of the synthesis observation.

The processing power in floating point operations per second (flop/s) required for convolution of the 4 polarized visibilities per baseline needed to keep up in real time with the output of the correlation process is analysed in subsection 3.7.1 and is for $\Delta\nu > \delta\nu$ given by

$$P_{\text{conv}} = 12 K_c^2 N_{\text{st}}^2 (\Delta\nu/\delta\nu) / \delta t \quad [\text{flop/s}] \quad (6.1)$$

where K_c^2 is the number of pixels in the 2-D convolution kernel, N_{st} is the number of stations $\Delta\nu$ is the total bandwidth, $\delta\nu$ is the channel bandwidth and δt the integration time per sample. The size of the complex convolution kernel that corrects for non-planarity as derived in subsection 3.4.2 is approximately given by

$$K_c \sim 14 \lambda H / D^2 \quad (6.2)$$

where H is the deviation from the reference plane by stations with diameter D .

The decorrelation of sources at half power of the station main beam at the longest baseline B_{max} is limited by choosing sufficiently small values for δt and $\delta\nu/\nu$ that are related to B_{max}/D as discussed in section 3.2 giving

$$P_{\text{conv}} = f_c K_c^2 N_{\text{st}}^2 (\Delta\nu/\nu) (B_{\text{max}}/D)^2 \quad [\text{flop/s}] \quad (6.3)$$

A factor 2-4 reduction of f_c can be obtained by increasing $\delta\nu$ and δt and accepting a reduced contribution by the longest baselines as discussed in subsection 3.4.6.

The processing power for an FFT per time interval Δt is

$$P_{\text{FFT}} = f_{\text{FFT}} (B_{\text{max}}/D)^2 \log(B_{\text{max}}/D) / \Delta t \quad [\text{flop/s}] \quad (6.4)$$

For multi-frequency imaging we can take the ratio of (6.3) over (6.4) where the resolution factor drops and P_{conv} will dominate over P_{FFT} if the time interval Δt is sufficiently long. This is the case for continuum imaging with LOFAR as shown in section 3.7. Subdividing a large field into a number of smaller fields gives a small decrease in processing power but complicates data administration.

When we insert (6.2) into (6.3) and take $H \sim B_{\text{max}}/2$ for conventional imaging with the plane of the Fourier image perpendicular to the line of sight (W -axis) towards the field centre we get

$$P_{\text{conv}} = 49 f_c N_{\text{st}}^2 (\Delta v/v) (\lambda/D)^2 (B_{\text{max}}/D)^4 \quad [\text{flop/s}] \quad (6.5)$$

This expression for single Fourier transform imaging using W -projection in the complex convolution of the visibility data, leads to acceptable processing powers only for short wavelengths and limited resolution but is not acceptable for LOFAR. We have shown in section 3.7 that full FoV imaging with Dutch LOFAR poses no processing problem in only two situations. One option is making a number of small facet images that effectively increase D by convolution of visibility data. The other option is working in the reference plane of the array using synthesized snapshot images with limited duration of about 10 min. The latter option has already been implemented and is called ‘ W -snapshots’ [Cornwell, Voronkov, Humphries, 2012], but it is not yet clear whether all aspects as discussed in chapter 3 have been properly appreciated.

In both approaches complex convolution is proposed and Dutch LOFAR would need ~ 500 facets for the LBA stations in their 32 m diameter configuration at 50 MHz when a minimum convolution kernel size of 7^2 pixels is used. Including the European stations providing the longest baselines but a narrower station beam about ~ 1800 facets would be required. For these large number of facets a fast faceting algorithm has been proposed in subsection 3.4.6 that could in principle be implemented on the correlation platform. Such an implementation would need a processing power that is a factor 1.7 greater than for correlation but allows reduction of the data output rate by selecting an appropriate subset of facets for imaging. For both imaging approaches we find processing power proportional to $(B_{\text{max}}/D)^2$ for FFT as well as for convolution, i.e. proportional to the FoV expressed in resolution elements.

Actual image forming needs additional processing to remove the artefacts by incomplete U, V -plane filling. We have to subtract a number of N_s strongest sources from the visibility data that is sufficiently large. The contribution by the side lobes of all remaining weaker sources to the rms noise in the image is then much weaker than the thermal noise. This requires additional processing power given by

$$P_{\text{sub}} = (f_s N_s / K_c^2) P_{\text{conv}} \quad (6.6)$$

The factor f_s has a value 2 - 3 and depends according to subsection 3.7.1.6 on the accuracy needed for subtraction of sources at large distance from the field centre.

According to subsection 5.2.6, the variance S_{rms}^2 of the flux of all sources weaker than S_s can be found by integrating the squared flux using the flux derivative of the integrated source density. The rms noise by the side lobes of all sources that are not subtracted gives a contribution

$$\Delta S = \epsilon_{\text{rms}} S_{\text{rms}}(<S_s) \Omega_{\text{mb}}^{1/2} \quad (6.7)$$

where Ω_{mb} is the solid angle of the main beam of the station and ϵ_{rms} the rms side lobe level within the beam. We require $\Delta S < 0.4 \Delta S_{\text{therm}}$ to increase the thermal noise ΔS_{therm} only by 8%. For a synthesis array with given thermal sensitivity and given ϵ_{rms} we can derive S_{rms} . An explicit expression for $S_{\text{rms}}(<S_s)$ as function of S_s can be derived from the integrated source density function $N(>S_s)$ given in table 4.2 and allows to find S_s . With known integrated source density function $N(>S_s)$ the total number of sources stronger than S_s can be determined that need to be subtracted, such that the remaining weaker ones contribute less than 8% to the thermal noise.

We have also shown that ϵ_{rms} is proportional to N_{st}^{-1} while Ω_{mb} is proportional to N_{st} , which means that for given ΔS we need according to (6.7) S_{rms} proportional to $N_{\text{st}}^{1/2}$. For the most relevant flux range we have S_{rms} proportional to $S_s^{1/2}$, which makes S_s proportional N_{st} . In that flux range the integrated source density is inversely proportional to S_s , which shows that an array with a larger number of smaller stations needs an equal amount of sources to be subtracted although the station beam is larger. At the high and at the low end of the flux regime discussed so far we get dramatic differences as demonstrated in table 5.1 for 35 MHz and for 140 MHz observing respectively, showing the importance of low side lobe levels.

These relations can also be applied to the noise that could be contributed by the psf side lobes of all sources in the sky attenuated by the side lobes of the stations and show that these contributions can be ignored if only the few strongest sources in the northern sky are subtracted. This requires accurate subtraction of Cas A, Cyg A, Tau A and Vir A and a few of the strongest sources that happen to be in a strong side lobe combination of an interferometer. It has been shown that these sources are also strong enough for proper self-calibration to allow such an accurate subtraction.

We compared various processing contributions in the image forming, but these should also be compared to other processing activity in a synthesis system. This is especially useful when this processing is executed on comparable platforms as is the case for LOFAR where both correlation of all station signals as well as image

forming run on multi-CPU platforms. The ratio of processing powers is in that case indicative for the cost ratio of the two processing activities. The correlation of a single spectral channel of bandwidth $\delta\nu$ [Hz] requires a processing power $\delta\nu$ [CMA/s] where a CMA is complex multiply add operation. Subtraction of a source needs 4 CMA for each visibility sample with integration time δt , which requires $4 N_s / \delta t$ [CMA/s] to keep up with the correlated output data stream, as discussed in section 3.7. The processing power ratio of subtraction over correlation is given by

$$P_{\text{sub}} / P_{\text{cor}} = 4 N_s / (\delta\nu \delta t) \quad (6.8)$$

This relation shows that for a typical Dutch LOFAR case with baselines shorter than 120 km using 10 kHz channels, 1 s sampling and subtraction of order 100 sources we need a platform for image forming that is only 4% of the correlation platform. However, subtraction of 1000 sources would increase this level to 40% which is in practice further increased by a factor ~ 2 if we account for multiple processing passes of the data. This limited increase is possible since smaller subsets get more passes to establish the proper parameters before the last pass handles all data. The situation becomes dramatic if full FoV imaging with 1200 km European baselines is considered. Then both spectral and temporal resolutions have to be increased by an order of magnitude. Full FoV imaging at full resolution using ~ 100 source subtractions per facet would require a processing platform far larger than the correlation platform if we need to keep up in real time.

6.2 Limitations by self-calibration

Chapter 4 discussed the effects of total electron content (TEC) in the ionosphere and more specific the effects of TEC gradients at scales spanning the extent of the synthesis array and at scales of the station beam extent at ionosphere height.

It has been shown that given the sensitivity of LOFAR, multi-direction self-calibration is possible even providing interpolated calibration parameters for sources between reference positions. Analysis of source count data and source size data has shown that such self-calibration is possible for the whole observing range of LOFAR, but the European stations need to use the baselines towards the LOFAR core. The accuracy of the interpolated phase parameters degrades with distance from the reference positions and first order estimates for this degradation have been derived using a Kolmogorov turbulence model that has been verified against observational results. In chapter 5 we discussed the impact of the degraded calibration parameters on the side lobes of the affected sources and their contribution to the effective noise in a synthesis image.

We have shown that the resolution provided by Dutch LOFAR allows estimation of the TEC from differential refraction by the curved ionosphere as function of frequency. This allows in principle self-calibration for Faraday rotation per 10 min interval from differential source positions in a synthesized snapshot image. Large

scale TEC gradients cause a constant position shift of the whole field as well as differential Faraday rotation that varies between different synthesized snapshot images. Differential TEC by travelling ionospheric disturbances (TIDs) also causes differential Faraday rotation but needs correction on time scales of order min related to the ionosphere coherence times for small scale Kolmogorov turbulence.

High quality imaging depends strongly on the stability of the ionosphere, but scintillation conditions that prevent imaging are not very frequent and occur mainly around the day/night terminators. More frequent are the TIDs that appear during varying fractions of most days. Fortunately, the medium scale TIDs that have sine wave like patterns with amplitudes up to 0.1 TECU and shortest wavelengths of about 90 km, are well sampled by the Dutch LOFAR array and its station beams.

LBA interferometers have sufficient sensitivity to observe at least 5 sources per beam, which allows determining a TEC value for 5 directions per station beam within an ionosphere coherence time using 20% relative bandwidth. With 5 reference positions a second order 2-D Lagrange polynomial can be determined, which could give an accurate description of a part of the TID if at most $1/6^{\text{th}}$ of a wavelength is sampled. This size also corrects for large scale Kolmogorov turbulence and corresponds to a station beam with an effective extent of $\sim 4^\circ$. The TEC distribution by the TID wave is in that case described with an accuracy comparable to deviations caused by residual small scale Kolmogorov turbulence. A larger beam, as provided by the LBA stations, allows an accurate interpolation only close to the reference positions, but further out the unpredictable Kolmogorov turbulence deviations could cause station based rms phase errors with respect to the reference positions that exceed 0.7 rad. Such phase errors would reduce the intensity of a point source by a factor smaller than 0.6 and create errors in the nominal side lobe pattern with the same rms value as of the nominal pattern.

Larger beams have overlap at the height of the ionosphere when stations are closer than a beam width at ionosphere height. This overlap provides additional sampling, which allows calibration of the shorter baselines with higher accuracy. The longer baselines provided by remote stations without overlap, suffer from the largest degradation in the visibilities of sources at larger distance from the reference calibrators. These longer baselines give therefore the largest contribution to side lobe noise, which could be reduced by additional tapering. In that case also the well calibrated sources will be broadened.

A fundamental limitation in multi-source self-calibration is that a solution for M directions per station needs at least M independent baselines from that station to other stations. The consequence for short narrow band snapshots with an array with N_{st} stations is a limitation of the total number of directions to $M < (N_{\text{st}} - 1)/2$. Larger relative bandwidth $\Delta\nu/\nu$ provides additional baseline samples on baselines B and station diameter D for which $D/B < \Delta\nu/\nu$.

A serious practical limitation is that all baselines shorter than 1 km have to be ignored and relative bandwidth larger than 9% is required to limit the contamination of the self-calibration solutions by all sources weaker than the weakest self-calibration source. The noise in the weakest self-calibration source is in that case no longer dominated by thermal noise for $M > 10$ which is a serious limitation since we need at least self-calibration and subtraction of the three strongest sources in the sky leaving ~ 7 self-calibration sources per station beam for interpolation at the remote stations. Fortunately Cas A, Cyg A and Tau A are resolved on the long baselines provided by the remote stations.

Station based phase errors in the visibilities smaller than 0.7 rad rms cause amplitude errors in the psf of a snapshot image proportional to the rms phase error. More importantly, the amplitude errors in the psf have an rms value that is also proportional to the square root of the psf of a snapshot image. HBA stations have indeed such small residual phase errors after self-calibration, which makes point source degradation negligible and contributions by error side lobes less than 7%. This is not only true for snapshot images, but integration over longer periods reduces these errors with the square root of time since these errors are independent after an ionosphere coherence time. The noise in a longer synthesis image reduces with the square root of time as well. The result is that also in a longer synthesis observation the ratio between the two components stays the same.

It has been shown that for a given bandwidth and integration time, HBA stations of different sizes provide about equal numbers of self-calibration sources per station beam. The reason is that larger stations have higher sensitivity that compensates to first order for the smaller station beam that has less chance to contain stronger sources. This has an important consequence for array design where a total number of antenna tiles have to be distributed over a large number of small stations or over fewer stations that are larger. The sensitivity of the tiles has to be such that remote stations with a beam width of 4° observe in an ionosphere coherence time at least 5 sources suitable for self-calibration. Smaller stations, which would give a better U,V-distribution, have a larger beam but need more sensitivity to observe more sources per beam to get the same source density. This allows higher order interpolation that gives full self-calibration for all sources in the beam in principle but could suffer from the limited number of source direction that can practically be solved for as discussed above.

6.3 System design of synthesis arrays

System design is different from system engineering, where the latter assumes a reference design that “only” needs to be detailed to a level where actual building can start. This distinction is crucial when new technology allows design of systems that only exist in imagination. Realization of a large system, however, for which

system engineering is an essential step, is within a context provided by society. Also design is constrained, often by lack of imagination, but foremost by fundamental laws of physics and by limitations set by current and forthcoming technology. Especially this last aspect where Moore's law predicts larger future processing performance at given cost, has been the key to success for realization of LOFAR and will be so for the SKA.

System engineering has many aspects of which planning is a crucial one that puts emphasis on using proven technology to reduce the risks of delay and budget overrun. This risk mitigation strategy often leads to less performance within a given budget, while the scientific user community favours performance, even if it is delayed. Especially new technology allows realization of systems that allow faster scientific progress within available budgets. The importance of Moore's law is that it allows system design based on not-yet-proven processing subsystems that will be available in due time.

The combination of imagination and forthcoming technology has led to the realization of LOFAR, a precursor instrument that has shown phased array and digital processing technology to work in practice. Successful calibration methods have been developed and this dissertation shows why these methods can work in the first place following reasoning from first principles and showing their ultimate limitations using first order approximations verified against experimental data.

The key aspect of system design of a synthesis array is the distribution of the total collecting area over a number of stations. Especially the phased array technology allows full flexibility with little cost impact. The total collecting area determines the ultimate sensitivity, but the distribution of the stations defines the U,V-distribution that determines confusion noise by limited resolution of the array as well as side lobe noise by incomplete coverage. A few but large stations provide only a small FoV and a psf with a high side lobe level due to limited instantaneous U,V-coverage. Multi-beaming would allow reuse of the collecting aperture by forming beams in different directions and could extend the instantaneous total FoV. A large number of smaller stations produce quadratic more baselines, while the FoV increases only linearly. This results in a quadratic larger correlation platform to support a required FoV, but also the platform for image forming scales at that same rate.

Reaching the thermal noise for continuum images using incomplete U,V-coverage needs subtraction of a large number of sources. Subtraction of more than 20 sources dominates over complex convolution processing with the smallest kernel, which in turn dominates over FFT processing for most applications. Interestingly, a multi-beam configuration with large stations needs the same number of source subtractions per beam as a configuration with more but smaller stations that have larger beams. Although the larger stations have a narrower beam the array psf has

a higher side lobe level and we need subtraction of weaker sources of which the number density is larger. As a result a processing efficient multi-beam configuration that fills a given total FoV needs subtraction of more sources at lower levels than a configuration with larger beams. However, the total processing is reduced since quadratic less baselines per beam need to be processed.

Design of an optimized system balances the cost of the different subsystems such that the marginal system performance over marginal cost ratio is equal for all subsystems. The performance metric could be maximum detection sensitivity or maximum survey sensitivity. The latter makes FoV an essential input parameter and system optimization shows that up to 50% of the total cost could be spent on items that increase the FoV in an attempt to improve the sensitivity of a survey that covers a larger field than the instantaneous FoV of a multi-beam system. In view of reducing cost of digital signal and data processing over time, it is however not attractive to spend a large initial investment on these items.

An important aspect of system design is that once solutions for all subsystems have been identified, further system engineering will identify alternative solutions that could even be more cost effective.

We have shown that large FoV and real time high resolution imaging, leads to very high processing power for image forming that could well exceed the processing power for correlation and is a serious concern for the SKA.

6.4 Recommendations for LOFAR

Processing requirement due to incomplete U,V-coverage could be mitigated by building a few additional stations. Such an investment needs however to be balanced against savings on processing platforms, that are substantial only if real time imaging is required.

The LOFAR LBA stations have wider beams than 4° , which means that full FoV self-calibration can only be realized in good ionosphere conditions

Multi-beaming with 7 wide beams allows observing the whole sky for $\delta > 0$ with only 20 observations of 6 hour that could be completed within a week. Since we need good ionosphere conditions, many trial observations are needed before actual imaging will be done. Then it is important that as much of the high quality data are stored asking for an adequate output data rate of the correlation platform.

The following recommendations are proposed

- In view of the output data rate of the correlation platform that is too low for wide field imaging with the European array, it is recommended to implement the fast-faceting method on the correlation platform, which allows selection of only relevant subsets from the total FoV for storage and further processing.
- Detailed suggestions for implementation of fast-faceting and complex convolution correction with a small kernel in existing packages are outside the scope of this dissertation.
- The recent implementation of “W-snapshots” that seems to follow the proposed synthesized snapshot imaging method needs to be verified against the limitations set out in chapter 3.

6.5 Recommendations for SKA-Low

Real systems will inevitably show effects in imaging results that are combinations of approximations used in subsystems. The focus in this dissertation has been on calibration and continuum imaging, more specific on their interrelation with array and station configuration, which determine required processing resources for sky noise limited performance.

An important design activity for the SKA is the evaluation of the results by LOFAR, which need careful tracing to subsystem performance to confirm whether they behave as planned. This is especially important for implementations that seek to average out expected residuals over time. This might require a large number of repeated observations that bring the thermal noise down and could reveal systematic effects that would show up in a single more sensitive SKA observation.

An important subject for detailed analysis is the configuration rotation of the stations to suppress noise contributions by all sources outside the station main beam that have only limited suppression by the side lobes of the station beam.

- Is the back rotation of the element antennas in a station essential for efficient aperture synthesis imaging?

Formally full U,V-coverage is needed for reliable imaging and limited U,V-coverage leads practically to reduced sensitivity by the noise introduced by the psf side lobes of all sources in the field. Self-calibration and accurate subtraction in the visibility domain can reduce this contribution, but if more than 20 sources have to be subtracted, this will dominate the processing power for image forming.

We have shown that processing for image forming with non-coplanar baselines can be minimized by using a combination of a small complex convolution kernel that regrid data to invoke the FFT for Fourier imaging. Conventional imaging with W-axis towards centre of the source field then requires potentially large numbers of facet images to fill the FoV defined by a station beam. We have presented an efficient approach for fast faceting that keeps the total data volume from the correlation operation constant.

- If this fast faceting would be implemented on the correlation platform, only the appropriate subsets for further processing could be selected, which reduces the output rate of the platform.

Synthesized snapshot imaging is an alternative approach that uses a coordinate system with W-axis towards local zenith, but third order phase terms limit the observing time to order 10 min.

- Synthesized snapshot imaging, as well as conventional facet imaging extended with fast-faceting and W-projection, provide scaling of processing for continuum imaging proportional to the total number of resolution elements in the total FoV, and this processing is dominated by source subtraction.

Full U,V-coverage can only provide low side lobes when the sample density is reweighted to get an uniform distribution that is appropriately tapered. Such a procedure will inevitable reduce the sensitivity of a Fourier image, but is attractive if the side lobe noise contribution can be brought down even further.

The recommendation is that

- Current weighting schemes are extended with one producing minimum side lobe level.

Multi-beaming is more processing efficient to obtain a large FoV than a larger number of small stations. However, more sources have to be subtracted to reach the thermal noise in continuum imaging since a configuration with fewer stations has a higher side lobe level. Not only more sources need to be subtracted, but more importantly these sources have lower fluxes.

This raises the important research question

- Can all sources that have to be subtracted can indeed be identified?.

A typical maximum instantaneous FoV is $\sim 200 \text{ deg}^2$ and a phased array station could be configured to allow such a FoV in principle. In practice we would install

processing equipment to support $\sim 100 \text{ deg}^2$, which allows a sky survey that covers at most $\sim 30,000 \text{ deg}^2$ to be completed with ~ 300 observations. Taking half a day as the basic observing unit, which defines the basic sensitivity in observed fields, we could repeat this sky survey for 6 years and improve the sensitivity by a factor 3.8. It is much more attractive to start a first instalment with less beams covering only 50 deg^2 and install platforms for correlation and image forming that cost only half. After 3 years we improved the survey sensitivity only by ~ 2.7 but the cheaper platforms could be then replaced at the same cost by 4 times more powerful ones. These new platforms handle 200 deg^2 and within $\frac{1}{4}$ of a year the survey sensitivity is raised to 3.8 leaving time for other observing while a more powerful system is available at the same total cost.

This last example shows that in view of growing performance over time of digital processing platforms, as indicated by Moore's law, we should adopt the principle that

- The investment in digital signal and data processing facilities, especially those that provide additional bandwidth or FoV should be spread over time.

6.6 Main results

The dissertation summarized in chapter 2 all new technologies and approaches used by LOFAR. As a pathfinder to SKA a solid body of "proven technology" has been established.

The main part of the dissertation, chapter 3, is a detailed analysis of aperture synthesis imaging, and a design study for processing efficient wide-field imaging showing that phased array stations are excellent elements in a low-frequency aperture synthesis array.

Chapter 4 summarizes the principles and limitation of wide-field self-calibration at low frequencies showing why it is possible at all and indicates how it could be implemented in principle.

The important side lobe noise is investigated in chapter 5 and it is shown that wide field self-calibration contributes an rms side lobe noise that is less than 35% of the thermal noise when the Traveling Ionospheric Disturbances are appropriately sampled by the station beam. Beams wider than 4° as for the LOFAR Low band will cause a larger increase than 7% of the thermal image noise. In addition, substantial fractions of the field-of-view (FoV) will show sources with reduced resolution.

The most important results are

- The processing for two proposed imaging methods scales for continuum imaging with the number of resolution elements in the imaged FoV (both are expressed in solid angle).
- Processing for image forming is for these methods dominated by the subtraction of sources if more than 20 sources have to be subtracted to reach an image noise close to the thermal noise.
- This subtraction processing is minimized when the side lobe level of the point spread function (psf) of a synthesis array is sufficiently low.
- A first order derivation for the rms in the psf of a configuration like LOFAR is presented that traced down which characteristic features that determine this rms level.
- A detailed analysis is given for the maximum beam width for stations in an aperture synthesis array that need wide-field self-calibration for thermal noise limited imaging at frequencies where the Traveling Ionospheric Disturbances dominate the phase errors between the stations.

LOFAR commissioning is still progressing and could show artefacts that might be larger than predicted by the first order approximations presented in this dissertation. Nevertheless all basic arguments have been presented that lead to fundamental limitations in imaging performance and these form a reference for further analysis.

Bibliography

1. Alexander, P., **Bregman**, J.D., Faulkner, A.J., "SKA Data Flow and Processing", Proc. SKADS Conf. 2009, Wide Field Science and Technology for the Square Kilometre Array, eds. S.A. Torchinsky, A. van Ardenne, T. van den Brink, A.J.J. van Es, A.J. Faulkner, ISBN 978-90-805434-5-4, 2010
2. Alliot, S., Soudani, M., **Bregman**, J.D., "Comparison of filters with poly-phase structure applied to large embedded systems for telescopes", Proc. IEEE Benelux Signal Processing Symposium, March 2002
3. Ardenne, A. van, Smolders, B. Hampson, G., "Adaptive antennas for radio astrometry: result from the R&D program on the Square Kilometer Array", Proc. SPIE, Radio Telescopes, Vol. 4015, pp. 420-433, 2000
4. Ardenne, A., van, "Concepts and Technical studies for the SKA", Proc. The Universe at Low Frequencies, (invited), IAU Symp. 199, Puna, India, Dec. 1999, Eds. A. Pramesh Rao, G. Swarup, and Gopal-Krishna, p. 467, 2002
5. Ardenne, A. van, "The technology challenge for the next generation radio telescopes", Perspectives on Radio Astronomy: Technologies for Large Antenna Arrays, Eds. A.B. Smolders and M.P. van Haarlem, pp. ix-xviii, ISBN 90-805434-2-X, 1999
6. Ardenne, A. van, Smits, F., "Technical aspects of SKA1", High sensitivity radio astronomy, Eds. Jackson, N., Davis, R.J., Cambridge press, ISBN 0 521 57350 5, 1997
7. Arts, M.J., "Low Band Antenna FTS-2 beam pattern simulations", LOFAR-ASTRON-RPT-064, 2006
8. Arts, M.J., "EM-simulations of the final design of the LOFAR low band antenna", LOFAR-ASTRON-RPT-107, 2005
9. Arts, M.J., "EM-simulations of a LOFAR low band antenna", LOFAR-ASTRON-RPT-026, 2003
10. Becker, R.H., White, R.L., Helfand, D.J., "The FIRST survey: Faint Images of the Radio Sky at Twenty centimeters", The Astrophysical Journal, 459, pp. 559-577, 1995
11. Bennett, A.S., "The preparation of the revised 3C catalogue of radio sources", Mon. Not. R. Astron. Soc. 125, pp. 75-86, 1962
12. Bernardi, G. et. al. "Foregrounds for observations of the cosmological 21 cm line: II. Westerbork observations of the fields around 3C196 and the North Celestial Pole", Astronomy & Astrophysics, 522, 2010
13. Boonstra, A.J., Tol, S. van der, "Spatial filtering of interfering signals at the initial Low Frequency Array (LOFAR) phased array test station", Radio Science, Vol. 40, 2005
14. Born, M. and Wolf, E., "Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light", Cambridge University Press, 7th ed., ISBN 0-521-64222-1, 1999

15. **Bregman**, J.D., "Dipole and Station rotation for LOFAR", LOFAR-ASTRON-RP, 2012
16. **Bregman**, J.D., "Scenario for SKA Processing Cost Development of Wide Field Imaging with Aperture Arrays versus Phased Array Feeds", SKA memo 131, Dec 2010
17. **Bregman**, J.D., "Towards a LOFAR Array and Station Configuration", LOFAR-ASTRON-MEM-030, 2005
18. **Bregman**, J.D., "System Optimisation of multi-beam Aperture Arrays for Survey Performance", *Experimental Astronomy* vol. 17 pp. 407-416, 2004a
19. **Bregman**, J.D., "Cost effective frequency ranges for multi-beam Dishes, Cylinders, Aperture Arrays and Hybrids", *Experimental Astronomy* vol. 17 pp. 365-380, 2004b
20. **Bregman**, J.D., Kant, G.W., Ou, H., "Multi-terabit routing in the LOFAR signal and data transport networks", *Proc. URSI-GA 2002*
21. **Bregman**, J.D. "Concept design for a Low Frequency Array", *Proc. SPIE, Radio Telescopes*, Vol. 4015, pp. 19-32, 2000a
22. **Bregman**, J.D., Tan, G.H., Cazemier, W., Craye C., "A wideband sparse fractal array antenna for low frequency radio astronomy", *Proc. IEEE symposium on Antennas and Propagation*, 2000b
23. **Bregman**, J.D., "Design Concepts for a Sky Noise Limited Low Frequency Array", *Perspectives on Radio Astronomy: Technologies for Large Antenna Arrays*, Eds. A.B. Smolders and M.P. van Haarlem, pp. 23-32, ISBN 90-805434-2-X, 1999
24. **Bregman**, J.D., "Design constraints for a sky noise limited low frequency wide field continuum imaging synthesis array", LOFAR-ASTRON-MEM-259, 2010, SKA Study Memo-0038, 1998
25. Brouw, W.N., "Some Thoughts on the SRT gridding procedures", NFRA note 186, Dec. 1974
26. Brouw, W.N., "Data Processing for the Westerbork Synthesis Radio Telescope", PhD Thesis University Leiden, 1971
27. Bunton, J.D., "A Method for the Inclusion of Square Law Phase Terms in Wide Field Mapping, private communication
28. Cappellen, W.A. van, Wijnholds, S.J., **Bregman**, J.D., "Sparse antenna array configurations in large aperture synthesis radio telescopes", *Proc. European Radar conf. Manchester*, Sept. 2006
29. Cappellen, W.A. van, **Bregman**, J.D., Arts, M.J., "Effective sensitivity of non-uniform phased array of short dipoles", *Experimental Astronomy* vol. 17 pp. 101-109, 2004
30. Carozzi, T.D., Woan, G., "A generalized measurement equation and van Cittert-Zernike theorem for wide-field radio astronomical interferometry", *Monthly Notices of Royal Astronomical Society*, Vol. 395, Issue 3, pp. 1558-1568, May 2009
31. Cohen, A.S., Lane, W.M., Cotton, W.D., Kassim, N.E., Lazio, T.J.W., Perley, R.A., Condon, J.J., and Erickson, W.C., "The VLA Low-Frequency Sky Survey, *The Astronomical Journal*, vol. 134, pp. 1245-1262, Sept. 2007

-
32. Condon, J. J., Cotton, W. D., Greisen, E. W., Yin, Q. F., Perley, R. A., Taylor, G. B., & Broderick, J. J., "The NRAO VLA Sky Survey", *Astronomical Journal*, 115, PP. 1693-1716, 1998
 33. Cornwell, T.J., Golap, K., Bhatnagar, S., 2008, "The Non-coplanar Baselines Effect in Radio Interferometry: The W-Projection Algorithm", *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 5, pp. 647-657, Oct. 2008
 34. Cornwell, T.J., "EVLA and SKA computing cost for wide field imaging", EVLA memo 77, ~2004
 35. Cotton, W., et. al., "Beyond the Isoplanatic Patch in the VLA Low Frequency Sky Survey", *Proc. SPIE* vol. 5489, pp. 180-189, 2004
 36. Ellingson, S.W., Cazemier, W., "Efficient multibeam synthesis with interference nulling for large arrays", *IEEE Transactions on Antennas and Propagation*, Vol. 51, issue 3, pp. 503-511, March 2003
 37. Falcke, H. et.al., "A very brief description of LOFAR – the Low Frequency Array", *Highlights of Astronomy*, Vol. 14 XXVth IAU General assembly, Aug. 2006
 38. Garrett, M.A., Bruyn, A.G. de, Giroletti, M., Baan, W.A. and Schilizzi, R.T., "WSRT observations of the Hubble Deep Field region", *Astron. Astrophys.* 361, L41-L44, 2000
 39. Gunst, A.W., Kant, G.W., "Signal transport and Processing at the LOFAR Remote Stations", *Proc. URSI-GA* 2005
 40. Haarlem, M.P. van, Wise, M.W., Gunst, A., builders list LOFAR, "LOFAR: The Low Frequency Array", to appear in *Astronomy & Astrophysics*, 2012
 41. Hamaker, J.P., "Understanding radio polarimetry IV. The full-coherency analogue of scalar self-calibration: Self-alignment, dynamic range and polarimetric fidelity", *Astron. Astrophys. Suppl. Ser.* 143, 515-534, 2000
 42. Hamaker, J.P., **Bregman, J.D.**, Sault, R.J., "Understanding radio polarimetry: I. Mathematical foundations", 1996, *Astr. & Astrophys. Suppl. Ser.* 117, 137-147
 43. Humphreys, B., Cornwell, T., "Analysis of Convolutional Resampling Algorithm Performance", SKA memo 132, Jan 2011
 44. Intema, H.T., "A sharp view on the low frequency radio sky", Ph. D. thesis, Leiden University, Leiden, The Netherlands, 2009a
 45. Intema, H.T., Tol, S. van der, Cotton, W.D., Cohen, A.S., Bemmell, I.M. van, Rötgering, H.J.A., "Ionospheric Calibration of Low Frequency Radio Interferometric Observations using the Peeling Scheme: I. Method Description and First Results", *Astr. & Astrophys.* 501, pp 1185, 2009b
 46. Ishwara-Chandra, C.H., Sirothia, S.K., Wadadekar, Y., Pal, S. and Windhorst, R., "Deep GMRT 150-MHz observations of the LBDS-Lynx region: ultrasteepest spectrum radio sources", *Mon. Not. R. Astron. Soc.* 405, pp 436-446, 2010
 47. Ivashina, M.V., Maaskant, R. Woestenburg, B., "Equivalent System Representation to Model the Beam Sensitivity of Receiving Antenna Arrays", *IEEE Antennas and Wireless Propagation Letters*, Vol. 7, 2008

48. Kassim, N.E., Lazio, T.J., Erickson, W.C., Crane, P.C., Perley, R.A., Hicks, B., "The Low-Frequency Array (LOFAR): Opening a New Window on the Universe", Proc. SPIE, Radio Telescopes, Vol. 4015, pp. 328-340, 2000
49. Kassim, N.E., Perley, R.A., Ericson, W.C., Dwarakanath, K.S., "Subarcminute resolution imaging of radio sources at 74 MHz with the Very Large Array", The Astronomical Journal, vol. 106, no. 6, pp. 2218-2228, 1993
50. Katgert, P., Oort, M.J.A., and Windhorst R.A., "The WSRT 1.4 GHz amalgamated source counts", Astron. Astrophys, 195, 21-24, 1988
51. Kraus, J.D., "Antennas", 2nd Edn., ISBN-0-07-100482-3, 1988
52. Kazemi, S., Yatawatta, S., Zaroubi, S., Bruyn, A.G. de, Koopmans, L.V.E, Noordam, J., "Radio Interferometric Calibration using the SAGE Algorithm", Mon. Not. R. Astron. Soc. 414, no. 2, p.1656, 2011
53. Lacy, M., Rawlings, S., Warner, P.J., "A complete sample of radio sources in the North Ecliptic Cap, selected at 38 MHz – I. The radio data", Mon. Not. R. Astron. Soc. 256, pp. 404-424, 1992
54. Lenc, E., Garrett, M.A., Tingay, S.J., "A deep high-resolution survey of the low-frequency radio sky", Astrophysical Journal, 673, pp 78-95, 2008
55. Lonsdale, C.J., "Configuration Considerations for Low Frequency Arrays", Astronomical Society of the Pacific Conference Series, vol. 345, pp 399+, Dec. 2005
56. Maat, D.H.P., Kant, G.W., "Fiber Optic Network Technology for Distributed Long Baseline Radio Telescopes", Experimental Astronomy vol. 17 pp. 213-220, 2004
57. McGilchrist, M.M., Baldwin, J.E., Riley, J.M., Titterton, D.J., Waldram, E.M., and Warner, P.J., "The 7C survey of radio sources at 151 MHz – two regions centred at RA 10^h 28^m, Dec. 41° and RA 06^h, Dec. 45°", Mon. Not. R. Astron. Soc. 246, pp. 110-122, 1990
58. Nieuwpoort, R.V. van, Romein, J.W., "Using many-Core Hardware to Correlate Radio Astronomy Signals", Proc. Int. Conf. Supercomputing (ICS'09), pp. 440-449, June, 2009
59. Nijboer, R.J., Pandey-Pommier, M., Bruyn, A.G. de, "LOFAR imaging capabilities and system sensitivity, SKA memo 113, 2009
60. Nijboer, R.J., Noordam, J.E., Yatawatta, S.B., "LOFAR Self-calibration using Local Sky Model", ADASS XV, Astr. Soc. Pacific Conference Series, Vol. 351, pp 291-294, ISBN 1-58381-29-9, 2006
61. Noordam, J.E., "LOFAR Calibration Framework", LOFAR-ASTRON-ADD-015, Oct. 2006
62. Noordam, J.E., "Self-calibration of radio-astronomical observations", Proc. SPIE, Radio Telescopes, Vol. 4015, pp. 366-369, 2000
63. Norden, M.J., **Bregman, J.D.**, "Lightning protection strategy used in LOFAR radio telescope," proc. EMC Europe 2010

-
64. Oort, M.J.A., Steemers, W.J.G., Windhorst, R.A., "A deep 92 cm survey of the Lynx area", *Astron. Astrophys. Suppl. Ser.* 73, pp 103-123, 1988
 65. Owen, F.N., Morrison, G.E., Klimek, M.D., and Greisen, E.W., "The Deep SWIRE Field. II. 90 cm Continuum Radio Observations and 20 cm – 90 cm Spectra", *The Astronomical Journal* 137, pp 4846-4853, 2009
 66. Owen, F.N., and Morrison, G.E., "The Deep SWIRE Field. I. 20 cm Continuum Radio Observations: A Crowded Sky", *The Astronomical Journal* 136, pp. 1889-1900, 2008
 67. Rao, Urvashi, "Parameterized Deconvolution for Wide-Band Radio Synthesis Imaging," Thesis New Mexico Institute of Mining and Technology, 2010
 68. Rees, N., "A deep 38-MHz radio survey of the area delta greater than + 60 degrees", *Mon. Not. R. Astron. Soc.* 244, pp. 233-246, 1990
 69. Rengelink, R.B., Tang, Y., Bruyn, A.G. de, Miley, G.J., Bremer, M.N., Röttgering, H.J.A., and Bremer, M.A.R., "The Westerbork Northern Sky Survey (WENSS) I. A 570 square degree Mini-Survey around the North Ecliptic Pole", *Astron. Astrophys. Suppl. Ser.* 124, 259-280, 1997
 70. Romein, J.W., Broekema, P.C., Mol, J.D., Nieuwpoort, R.V., "The LOFAR Correlator: Implementation and Performance Analysis", *Proc. ACM Symposium PPOPP'10*, pp. 169-178, Jan. 2010
 71. Romein, J.W., Broekema, C.P., Meijeren, E. van, Schaaf, K. van der, Zwart, W.H., "Astronomical Real-time Streaming Signal Processing on a Blue Gene /L Supercomputer", *Proc. ACM Symposium on Parallel Algorithms and Architectures*, pp. 59-60, July 2006
 72. Röttgering H.J.A. et al. (**Bregman**), "LOFAR – Opening up a new window on the Universe," proceedings of the conference "Cosmology, galaxy formation and astroparticle physics on the pathway to the SKA", Oxford, April 10-12, 2006
 73. Sault, R.J., Staveley-Smith, L., Brouw, W.N., "An approach to interferometric mosaicing", *Astronomy & Astrophysics Supplement Series*, vol. 120, pp. 375-384, 1996
 74. Schaaf, K. van der, Broekema, C.P., Diepen, G. van, Meijeren, E. van, "The LOFAR Central Processing Facility Architecture", *Experimental Astronomy*, Vol. 17, pp 43-58, 2004
 75. Schaaf, K. van der, **Bregman**, J.D., Vos, C.M. de, Achterop, S., Spaanenburg, L., "Hybrid Cluster Computing and Software in the LOFAR Radio Telescope", *Proc. Int. Conf. Parallel and Distributed Processing Techniques and Applications*, pp. 695-701, 2003
 76. Smirnov, O.M., "Revisiting the radio interferometer measurement equation I. A full-sky Jones formalism", *Astronomy & Astrophysics*, 527, A106, 2011
 77. Smirnov, O.M., Noordam, J.E., "MeqParm: Parameter Handling in the MeqTree System", *Proc. ADASS XV, APS Conference series*, Vol. 351, 2006
 78. Spoelstra, T.A.Th, "A climatology of quiet/disturbed ionosphere conditions derived from 22 years of Westerbork interferometer observations", *Journal of Atmospheric and Terrestrial Physics*, vol. 58, no. 11, pp. 1229-1258, Aug. 1996

79. Tan, G.H., Rohner, Ch., "The Low Frequency Array active antenna system", Proc. SPIE, Radio Telescopes, Vol. 4015, pp. 446-457, 2000
80. Taylor, G.B., Carilli, C.L., and Perley R.A., eds., "Synthesis Imaging in Radio Astronomy II", ISBN 1-58381-005-6, A.S.P. Conference Series, 1999
81. Thompson, A.R., Bregman, J.D., "System considerations for the SKA", Tech. Rep. Memo 74, SKA memo series, 2006
82. Thompson, A.R., Moran, J.M., Swenson Jr., G.W., 2001, "Interferometry and Synthesis in Radio Astronomy", 2nd edition, ISBN-13: 978-0-471-25492-8, WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, 2004
83. Tol, Sebastiaan van der, Rottgering, Huub, "Ionospheric Measurements with the Low Frequency Array (LOFAR)", proc. of the 13th International Ionospheric Effects Symposium, May, 2011
84. Tol, Sebastiaan van der, "Bayesian Estimation for Ionospheric Calibration in Radio Astronomy", Thesis, Delft, 2009
85. Tol, S. van der, Jeffs, B.D., Veen, A.J. van der, "Self Calibration for the LOFAR Radio Astronomical Array", IEEE Trans. On Signal Processing, vol. 55, no. 9, pp. 4497-4510, Sept. 2007
86. Veen, Alle-Jan van der, Leshem, A., Boonstra, A.J., "Array Signal Processing for Radio Astronomy", Experimental Astronomy, Vol. 17, pp. 231-249, 2004
87. Velthoven, P.F.J. van, "Medium scale Irregularities in the Ionospheric Electron Content", Ph. D thesis University of Eindhoven, 1990, Eindhoven, The Netherlands
88. Vos, C.M., Gunst, A.W., Nijboer, R.J., "The LOFAR Telescope: System Architecture and Signal Processing", Proceedings of the IEEE, Vol. 97, no. 8 Aug. 2009
89. White, R.L., Becker, R.H., Helfland, D.J., and Gregg, M.D., "A catalogue of 1.4 GHz radio sources from the FIRST survey", The Astrophysical Journal, 475, pp. 479-493, 1997
90. Windhorst, R. A., Mathis, D., & Neuschaefer, L., "The Evolution of the Universe of Galaxies: Hubble Centennial Symposium", ASP Conf. Ser. 10, ed. R. G. Kron, pp. 389-403, 1990
91. Wijnholds, S.J., **Bregman**, J.D., Ardenne, A. van, "Calibratability and Its Impact on Configuration Design for the LOFAR and SKA Phased Array Radio Telescopes", Radio Sci., 46, RS0F07, doi:10.1029/2011RS004733, 2011.
92. Wijnholds, S.J., "Fish-Eye Observing with Phased Array Radio Telescopes," Thesis University Delft, ISBN 978-90-9025180-6, 2010
93. Wijnholds, S.J., "Mutual Coupling, inter-tile Spacing and interstation Rotation in the HBA tile", LOFAR-ASTRON-MEM-245, ASTRON, 2008
94. Wijnholds, S.J., "Study on LOFAR Core Calibratability Based on Source Statistics", LOFAR-ASTRON-MEM-205, ASTRON, 2006
95. Wijnholds, S.J., "Confusion Limited All Sky Imaging with LOFAR's Initial Test Station Applying Wide Field Calibration Techniques", Proc. URSI-GA, 2005

-
96. Wijnholds, S.J., **Bregman**, J.D., Boonstra, A.J., "Sky Noise Limited Snapshot Imaging in the presence of RFI with LOFAR's Initial Test Station", *Experimental Astronomy*, Vol. 17, pp. 35-42, 2004
 97. Yashar, M., Kembell, A., "Computational cost of radio imaging algorithms dealing with the non-coplanar baseline effect: I ", TDP-CPG-memo 17, 2009
 98. Yatawatta, S., Zaroubi, S., Bruyn, G. de, Koopmans, L., Noordam, J., "Radio Interferometric Calibration using the SAGE Algorithm", in *IEEE 13th Digital Signal Processing Workshop and 5th IEEE Signal Processing Education Workshop*, pp. 150-155, 2009
 99. Yatawatta, S., "On the interpolation of calibration solutions obtained in radio interferometry", *MNRAS*, in press 2012a
 100. Yatawatta, S., Bruyn, A.G de., members EoR group Groningen, "Initial (deep) LOFAR observations of Epoch of Reionization windows: I. The North Celestial Pole", *Astronomy & Astrophysics*, submitted Oct 2012b

Summary

An important question about a dissertation is its classification. The sheer size of this book places it between a textbook and a collection of papers aimed at a specialist audience. It deals with wide-field Fourier imaging, and it shows from first principles how the burden of pre-processing interferometer data can be minimized. In a step-by-step explanation, the non-specialist is guided along the approximations that are necessary for efficient processing, and towards the scaling laws that govern wide-field continuum imaging.

The design of a new instrument brings together specialist information from many disciplines. It starts with an analysis of comparable instruments, in particular *how* and *why* they work. (The *what* is the third dimension of the knowledge volume, and will concern us later). This thesis starts with showing why LOFAR can be self-calibrated when its stations consist of a sufficient number of element antennas, and its synthesis array consists of a sufficient number of stations.

The available body of scientific knowledge turned out to be a kind of Swiss cheese, i.e. it has some structure that can stand scientific scrutiny, but also has many voids. Most of the spanned volume does not contain substance that is suitable as input for system design. For instance, consider the engineering paradigm that states that a large system should only use proven technology in all its subsystems. This is a recipe for slowing down progress to a pace set by the progress in understanding an by partial implementation. Although the use of this paradigm may avoid showstoppers when building a new instrument, the stumbling blocks that really need scientific attention are only found when running into them. In reality, most things that work well are not yet fully understood at all. Although science has many analytic tools, it lacks the synthesis tools that are needed for construction. A simple example is a butcher who exactly knows how to dissect a pig in its different parts, but is unable to put them together again into an entire animal, let alone a living one. Being alive is the essence of a working entity, and the only known way of creating life is from life, i.e. by taking living elements and let them grow together organically, making their way around any stumbling blocks on the projected shortcut road towards a not very well-defined destination.

This organic engineering synthesis process is reflected by this dissertation. It takes large steps over a broad front where scientific proof and full understanding exist, but much smaller steps in specific areas where new bridges must be defined to take us to the desired images with less effort. Although the focus is on *why* low frequency synthesis imaging can work at all, the *how* is also presented in order to reassure a new generation that the next milestone is indeed within reach. Many detail are not yet covered, preventing blindly surging ahead, but by going forward deliberately

while being sensitive to, and aware of, problems that must (and can) be solved on the way.

Therefore, a key issue in system design is to identify any *fundamental limitations*, and to learn how comparable systems deal with them. Practical solutions are often not driven by fundamental limitations, but by problems associated with premature choices made at concept design level. These choices are “premature” because they are driven by the knowledge and technology that is available at the time, but could well be obsolete by the time of final realization.

New elements in the concept design of LOFAR

During the 1990's, aperture synthesis observations at a frequency of 74 MHz with the 27 antennas of the VLA showed that the sensitivity was not sufficient for the self-calibration that was needed to correct for ionosphere disturbances of baselines up to 30 km. Even more serious was the large field-of-view of the 25 m dish antennas that are too small compared with wavelength. This violated the assumptions of the existing imaging software packages, which were developed for narrower antenna beams at shorter wavelengths.

The most important conclusion from this for future array design was the need for enough bright sources in the field to sample the shape of the station voltage beams with the help of multi-direction self-calibration. This then translated directly to the requirement for larger stations with adequate sensitivity.

By the turn of the century, digital signal processing equipment became available that made phased-array stations much larger than 25 m affordable for astronomical use. In addition, because of their electronically steered beams, they do not require mechanical tracking to observe celestial objects from a rotating Earth. Even more importantly, the well-known Moore's Law predicted a quadrupling every three years of the performance of digital processing, allowing adequate performance within an affordable budget for a large low-frequency array by 2003. Relying on the timely availability of adequate components and processing platforms was the basis of the concept design for LOFAR that was presented in 1999.

The actual design of digital receiver systems and calibration and imaging software could now start right away, based on preliminary specifications provided by industry for future components and processing platforms. A more extensive overview of all the new aspects in LOFAR is presented in chapter 2. The single most important technology breakthrough was the announcement of low-cost gigabit transceiver technology for data transport over optical fibre. This rendered it affordable to consider stations that were located up to hundreds of kilometres away from the central processor.

The LOFAR initial test station was completed in 2003. It used the first generation of new components, demonstrating wide-band short-dipole antennas, digital receivers, cross-correlation on a cluster of processors, and self-calibration for antennas at station level.

The flat station array allowed imaging covering a full hemisphere, which provided identification of the exact direction of various signal sources and of moving sources in a snapshot image. Combining a set of such snapshot images provided a large sky image covering more than a hemisphere.

Efficient imaging approaches

This pioneering effort showed the way forward for analysis of the limitations in 2-D Fourier imaging given in chapter 3. This analysis resulted in proposing two new imaging methods based on a different combination of demonstrated techniques that deal with field-of-view (FoV) limitations by non-coplanar baselines in 2-D Fourier imaging.

The so-called *faceting* technique reprocesses the whole set of interferometer data for each facet image to obtain a large number of small facet images to cover the field of a large station beam. The so-called *W-projection* method uses a complex quasi-convolution to correct the data of non-coplanar baseline before Fourier transformation.

Both methods require too much processing power to be of practical use for LOFAR, which begged for an analysis of the whole imaging process. The analysis revealed that both approaches deal with so-called *extrinsic* non-planarity by projection of baselines on the direction of the field-of-view. These large and varying projections are the result of Earth rotation when a telescope tracks a point in the sky. An important feature of both methods is the simplified correction for rotation of the baselines.

For a large FoV, it is attractive to use an imaging method that uses the much smaller *intrinsic* non-planarity of a large synthesis array of which the stations follow Earth curvature.

An analytic analysis of the complex quasi-convolution correction method revealed its dependence on non-planarity and FoV, showing two alternatives for efficient imaging that require the same minimum amount of convolution processing.

The first method follows the conventional imaging approach with extrinsic non-planarity, but uses larger facets allowed by a limited convolution correction. The method is particularly attractive for baselines longer than a few hundred kilometres and minimizes the required processing with a new fast *faceting* technique. This fast faceting technique is particularly powerful for continuum observations as with

LOFAR, and is based on a butterfly technique comparable to the one used in Fast Fourier transformation. The set of interferometer data is reorganized into a number of subsets that together have the same data volume. The important feature is that each small subset needs only a small Fourier transform for each small facet image. As a result, the total data processing is equal to the processing for a single large image, but also the distortions due to non-planarity (intrinsic and extrinsic) are almost fully corrected.

The second method is based on individual snapshot images made with a 2-D Fourier transform that could cover a hemisphere around Zenith for a strictly planar array. We avoid the large *extrinsic* non-planarity caused by projection effects, and need only correction for the much smaller *intrinsic* non-planarity of the array itself. An important advantage of this approach is that it clarifies in a straightforward manner how the imaging accuracy degrades for objects all over the sky. We suffer from such degradation by non-planarity but also by rotation corrections that are only valid for the centre of the field tracked by the station beam.

A limitation of the method is that the non-planarity also limits the maximum duration of a synthesized snapshot image to order 10 min for observations with stations up to 90 km from the centre of the array. For each synthesized snapshot image, the interferometer data need correction for a small shift and a small rotation of the tracked sky field before transformation. Simple first order correction leaves residual errors comparable with residual non-planarity errors.

The synthesized snapshot images need corrections before they can be combined to a single sky image. An image scale correction is needed since the projection of the sky is different for each image, while the parallactic rotation varies over the FoV. Next to these position corrections every 10 min, we also need intensity corrections. The images for the 4 polarizations need to be combined and corrected for instrumental polarization and parallactic rotation to give an image in each of the 4 Stokes parameters. The corrections per image pixel change only gradually over the FoV.

A synthesis observation longer than about 10 min therefore needs a number of large Fourier transforms, which is for LOFAR and arrays with more stations no longer the dominating processing.

In both methods, the amount of processing for image-forming is proportional to the number of resolution elements in the image, i.e. the solid angle of the station FoV, expressed in the angular resolution of the array as a whole. Implementing these approaches on a scale appropriate for LOFAR is in progress, forming a basis that is particularly suitable for an even larger instrument like the SKA.

System design and system engineering

The Square Km Array (SKA) will be developed and built in a number of stages. System engineering for the proposed SKA-low instrument asks for proven technology to estimate the size of platforms suitable for signal and data processing. Scaling the performance of existing imaging packages, which are designed and optimized for much smaller instruments, produce the rather unsatisfactory result that the processing platform needed for image-forming could easily absorb half the budget of a large imaging array. This raises the question whether we are just dealing with sub-optimal software design, or limited by the fundamental *scaling laws* that govern the processing for wide-field imaging at low frequencies.

System design concerns itself with combining subsystems in such a way that a final goal is realized for minimum cost. System design for scientific research starts from a given budget and asks for maximum return on investment. It is customary for scientists to define the scientific goals of a new instrument, after which engineers design it and calculate the required budget. However, relying on proven technology may lead to predictable cost and time scales, but could easily offer outdated performance when finished.

System engineering concerns itself with separating a large system into a set of subsystems that can be designed independent of each other. This is important to work concurrently with a number of design teams, where each team has its own set of specialised engineers. Within the field of electronics we have antenna, receivers and digital engineers and we need appropriate interfaces to transfer signals from one domain field to another domain field. Often we have a transmission line between an antenna and a receiver. Conventionally, antenna engineer match an antenna to the transmission line and the receiver engineer match the transmission line to the low noise transistor.

System design concerns itself with defining such interfaces in a way that minimizes cost, even allowing results that cannot be obtained with conventional approaches.

A striking example of questioning received wisdom is the short dipole, known to antenna engineers as a narrow-band device. But if we discard the underlying engineering paradigm that insists on power matching (needed for transmission, which is not relevant for a receiving instrument), it turns out to be a wide-band element that can be sky-noise-limited over two octaves in the frequency range where LOFAR operates.

For low-frequency sky-imaging, the phased-array station based on short dipoles has been identified as a building block that satisfies the requirements: a given budget for a given total number of element antenna defines the system sensitivity, which can then be distributed over a number of stations with little impact on cost. A large number of small stations results in good aperture sampling and a large field-

of-view (FoV), while an array with fewer large stations requires less processing but more station beams to cover the same FoV on the sky.

The best choice for a configuration depends on the application, but this dissertation concentrates on the most sensitive application, which is also the most demanding in terms of processing: wide-field imaging using a large relative bandwidth.

The most important questions are (i) how the configuration and number of stations determine the non-thermal side-lobe noise and the associated processing, and (ii) how station size limits the self-calibration performance of a synthesis array and introduces additional non-thermal noise. Chapter 3 discusses the minimum amount of processing that is required for Fourier imaging, while chapter 4 concludes that a minimum station size is defined by the scale size of ionospheric structure. Chapter 5 discusses how the artefacts caused by imperfect calibration and a limited number of stations determine the effective sensitivity, which is the primary cost driver for a synthesis array. The results of these three chapters are combined in chapter 6 to offer conclusions and recommendations for system design and further research.

Array configuration and side-lobe noise

In any imaging instrument, the image of a point source is convolved with a point-spread-function (PSF), the shape of which is determined by the sampling of the aperture plane. In radio aperture synthesis instruments, the sampling is relatively sparse, causing a PSF with considerable side-lobes, which extend over the entire FoV. Thus, the PSF of a bright source will effectively drown out fainter sources, limiting the dynamic range of the observation. Therefore, the brightest sources must be identified first, and their contributions subtracted from the observed visibility data *before* transforming them into an image. A Fourier Transform of the *residual* data will then produce an image that shows the fainter sources. For wide-band continuum observing with arrays like LOFAR, the source subtraction operation will dominate the processing for image forming if more than 20 sources have to be subtracted. This subject is extensively discussed in chapter 3.

Therefore, a very important question for system design is how many sources have to be typically subtracted, and how that number is influenced by the array configuration? This subject is discussed in chapter 5. Since the thermal noise level in a Fourier image is increased by the average of the PSF side-lobes of all the objects in the field, it is important to minimize the side-lobes of the PSF. This may be achieved by improving the sampling of the aperture plane. For instance, by using an array with more and/or better-placed stations, or by using a wider relative bandwidth, or by observing longer while the Earth rotates. In addition, side lobes may also be reduced considerably by applying baseline-dependent weights to the visibility samples when transforming them to an image. (NB: The PSF side lobes will also be affected by calibration errors, which are ignored here).

The auto-correlation of the station distribution defines the sampling function of the aperture plane. In chapter 5 we introduce a simplified configuration model for LOFAR, where half the stations are concentrated in a central cluster, and the other half are distributed over annuli with radii that increase exponentially. Such a configuration may be optimized for continuum imaging by adjusting its three characterizing parameters: the diameter of a station, the diameter of the core cluster, and the diameter of the entire array. The ratio of station diameter over array diameter defines a characteristic bandwidth relative to the observing frequency and a characteristic observing time that provide good aperture sampling over an aperture area with a station diameter.

The rms side-lobe level of the PSF with which the sources in a single snapshot image will be convolved is primarily determined by the number of stations in the array. It decays with the distance from the PSF centre, i.e. the position of a source. This decay is a complicated function, but far from the centre it is inversely proportional to the distance. With the above-mentioned characteristic relative bandwidth and snapshot observing time, this side-lobe decay sets in at the half-power distance of the station beam, for a source in the centre of the field.

We can increase the observing time and the relative bandwidth in a multi-frequency synthesis until they equal the ratio of the station diameter over the average distance between stations in the central cluster. The aperture is then filled with a pattern of clusters of independent visibility samples, which leads to a reduction of the radius where the side lobe decay sets in. A further increase of the observing time for a synthesized snapshot image only has a limited effect, since we need to fill the aperture by more but independent baseline clusters, and not with more samples in each cluster.

The maximum number of independent additional clusters is defined by the ratio of the array diameter over the cluster diameter. It shows the importance of distributing additional observing time and additional bandwidth in such a way that independent clusters of visibility samples are formed. The important result is that the side-lobe noise in a synthesis observation scales for short observing time and limited bandwidth, at the same rate as the thermal noise. However, for longer time and larger bandwidth this is no longer true.

Evaluation of the rms side lobe level for the LOFAR low-band array, using this simplified PSF model, shows that the noise contribution by the side lobe of all sources outside the main lobe of the station beam is less than 3% of the thermal noise, contributing less than 0.05% to the image noise. The main reason for this low contribution is that the station side lobes effectively suppress all outlying sources, except for a few very bright ones that just reach a level where they can be self-calibrated, and subsequently, be subtracted properly from the visibility data.

It is a property of the intensity distribution of the sky sources that sources weaker than the 4 strongest in the Northern hemisphere are one to two magnitudes weaker and therefore need no separate subtraction.

Assuming a typical PSF side-lobe distribution, at least the 100 strongest sources in the station beam have to be subtracted, in a 12 hour observation with a 40 station array, using 1% relative bandwidth @ 35 MHz. After subtraction, the side-lobe noise of all remaining sources is less than 40% to the thermal noise. This contribution adds at most 8% to the thermal noise in an image, but can be reduced by subtracting more sources. This evaluation assumes perfect subtraction, which is only possible for the few strongest sources in a station main beam and the few strongest in the rest of the sky that are individually self-calibrated.

Self-calibration and configuration impact

Traditionally, self-calibration only solves for a single complex gain error per station, assuming it to be valid over the entire field of view. In generalized (third-generation) calibration, it is recognized that some instrumental effects are direction-dependent, so it is necessary to solve for more parameters per station. The power of multi-direction self-calibration is explained to non-specialists in chapter 4. Among other things, it discusses the limitations of interpolating calibration parameters that vary rapidly in time, frequency or position. For instance small-scale ionosphere instability, which varies at time-scales shorter than about 1 minute. This leads to constraints on the size and sensitivity of the stations.

The explanation is derived from first principles, and places the small-scale disturbances in the framework that describes the time-dependent refraction caused by large-scale ionosphere structure. In principle, these large-scale terms vary slowly and can also be self-calibrated. Separation from faster small-scale effects is possible by averaging over intervals of order 10 min, a typical value for a synthesized snapshot.

It turns out that self-calibration can only solve for a limited number of parameters per station. This number is fundamentally limited by the number of independent baselines in which each station participates, but in practice it is limited by the noise in the measured visibilities. Iterative solving algorithms usually deal with parameter solving per source direction in order of decreasing source intensity, down to a flux of about three times the effective noise per visibility. This effective noise contains a contribution by the flux of all sources that are too faint to be solved for, and defines a maximum number of sources that can be solved for.

We derive a first-order estimate, based on the actual density of sources, as a function of flux, showing that this number is about 10 when the noise contribution by all contaminating sources equals the thermal noise. Moreover, a relative bandwidth

larger than 10% is needed in a snapshot dataset with the duration of an ionosphere coherence time. Only baselines longer than a km should be used in the solution, so as to reduce the contributions of contaminating sources.

An additional constraint is that typically ~ 4 very bright sources outside the station beam (i.e. the so-called A-team: Cas A, Cynus A, etc) have to be solved for and subtracted as well. The important result is that at least 5 sources (directions) per station can be solved for, which is adequate to model the phase errors over the station beam by medium-scale travelling ionospheric disturbances (TID), provided that beam is smaller than about 4° .

LOFAR has enough stations of different sizes to provide sufficient sensitivity on a number of baselines, allowing self-calibration and high quality full FoV imaging in the high frequency band (115 - 230 MHz). Wide-field imaging in the low frequency band (10 - 90 MHz) is complicated by relatively large gaps between the self-calibration sources caused by large station beams and limited sensitivity. As a result, the interpolated phases do not represent the actual ones of the TIDs in the ionosphere. The magnitude of the phase errors increases with observing wavelength and separation between the sources used for self-calibration, and allows high quality imaging only for a limited fraction of the station beam.

An important question is how the source flux that is scattered by phase noise in the visibilities propagates into the Fourier image. Especially, whether this increases the thermal noise above the level determined by receiver noise, global sky brightness, bandwidth and observing time.

All sources that use interpolated calibration parameters suffer from errors that increase with their distance from the self-calibration sources (where the errors are assumed to be zero). These interpolation errors lead to a distortion of the PSF with which each source is convolved, causing additional noise by the errors in the side lobes of these sources. A first-order estimate using the phase noise in the interpolated calibration parameters suggests that the noise, left after subtraction using interpolated calibration parameters, is less than 38% of the thermal noise. This will add at most 7% to the thermal noise in an image, but it cannot be reduced any further. This is a generic result, which is valid for the noise introduced by all sources that could not be subtracted accurately using their own self-calibration parameters, but had to use interpolated parameters.

In addition to these interpolated self-calibration contributions, there are phase errors due to Kolmogorov turbulence in the ionosphere, which also increase with distance from the self-calibration sources. For the wide main beam of the LOFAR low band stations, which have large gaps between the self-calibration sources, the phase errors per coherence time can reach an rms value larger than 0.7 rad per station. On shorter baselines, stations are sufficiently close together to share calibration information, so that interpolated phase errors will be smaller. The result is that, at a

given location in the station beam, the phase errors increase beyond 1 radian per baseline.

For such large phase errors, the distortions of the PSF side-lobes can no longer be described as perturbations of the nominal PSF. However, the resulting PSF has a side-lobe distribution with the same rms value. An important aspect is that the main lobe of the PSF will also break up, which will result in a so-called speckle pattern. In this case, the averaging of many speckled snapshots will produce severe blurring of sources in a substantial fraction of the main beam. This will not only reduce the peak intensity of objects in these areas, but also give additional error side lobes that contain the scattered power from these sources. A first order estimate of this noise contribution, dominated by the longest baselines, is less than 45% of the thermal noise. This contribution can be reduced by ignoring these baselines, but this will also reduce the resolution of all sources near the self-calibration sources.

Finally, we need to combine the various error contributions by adding their squared rms values. This yields an increase of at least 7% in the image noise for LOFAR high band observations, and typically 23% for the low band.

Design optimization and processing scaling

For appropriate system design, we need the scaling laws that determine the optimum distribution of cost over subsystems for a system with an imaging quality and effective sensitivity that is matched to the nominal sensitivity provided by total collecting area and by the signal bandwidth. We have indicated in the previous paragraphs that the final noise in an image is not only determined by the thermal sensitivity of the instrument, but also by its configuration. The derived dependencies allow us to compare the impact of alternative system configurations on the total system performance.

A station size that is too small, will offer limited calibration accuracy, which will introduce additional noise that cannot be recovered. Too few stations will cause a high PSF side-lobe level that will require additional processing for source subtraction. For example, a 10% sensitivity loss by insufficient processing can be reduced to 5% by increasing the imaging platform at large additional cost to subtract more sources, which is however only effective for continuum observing. Alternatively, the number of stations could be increased by 5%, which has a serious cost penalty, and is not needed for spectral line observing. Also, the total FoV could be increased by forming 10% more beams per station, but this helps only in survey applications.

In chapter 3 we conclude that the minimum processing power for real time continuum imaging is proportional to the solid angle of the FoV, measured in resolution elements, i.e. proportional to the square of the ratio of the array diameter over the station diameter. It is also shown that this processing is dominated by source sub-

traction if more than 20 sources have to be subtracted. Chapter 6 combines this result with the results of chapter 5 and provides global scaling laws for signal and data processing that will be summarized in the following paragraphs.

Phased-array technology allows flexible distribution of the total affordable collecting area over a number of stations, which may even have different sizes, while the FoV can be controlled by forming more beams per station. Given a minimum station size, the total number of stations is defined by the available budget.

Alternatively, a configuration with fewer but larger stations and more beams could in principle provide the same sensitivity and the same total FoV. The latter configuration has a smaller number of baselines, while the required number of beams per station is increased. Although the total input bandwidth of the correlation platform is the same in that case, the totally required processing power decreases linearly with the number of stations. Less obviously, the output sample rate for continuum imaging is reduced at the same rate, which is an attractive feature for an imaging platform that needs to handle the correlated data in real time. From the perspective of correlation platform design, the multi-beam solution looks preferable.

Our analysis has shown that a configuration with fewer stations measures quadratic less visibility samples, which leads to a reduced image quality. Although the station beam is narrower, at least the same number of sources have to be subtracted. The same number of sources in a smaller beam just means that weaker sources have to be subtracted to reach the same level of image noise as with a configuration with more but smaller stations. These stations have a wider beam that is less sensitive, but detects equal numbers of self-calibration sources. Consequently, the distance between the self-calibration sources increases, and calibration quality after interpolation degrades. This leads to additional noise in an image.

The processing for source subtraction is proportional to the number of visibilities, and dominates continuum synthesis imaging. In a configuration with larger stations and consequently more beams, we have to subtract at least proportionally more sources. The number of baselines is however reduced stronger, and consequently, the total processing for continuum image forming is reduced just like the processing for correlation of all telescope signals.

We have however shown for the high band case of LOFAR that the number of sources that have to be subtracted increases progressively when the fluxes of these sources reach the thermal noise level. In that case, the potential processing advantage for imaging with an array using less but larger stations becomes questionable.

Further study needed on the minimum number of stations

It seems attractive to start with an array with relatively few large stations, of which the effective FoV can be enhanced in a later stage by forming additional station beams by means of more processing. For continuum imaging, such a choice has a side lobe noise that is determined by the number of stations and their configuration. It could be well above the thermal noise, especially when observations are repeated many times to reduce the thermal noise, while the side lobe noise is identical in each repeated observation. This side-lobe noise can only be reduced by enhancing the configuration with more stations.

Optimization of the configuration of a synthesis array with phased array stations depends therefore strongly on the ultimate sensitivity that needs to be reached.

Therefore, an important subject for detailed further analysis is whether the additional and fainter sources that have to be subtracted in a multiple narrow beam configuration can be identified at all, since they are much closer to the noise floor. If that is not the case, a higher final noise-level has to be accepted for continuum imaging, which cannot be reduced when more processing power becomes available in a later stage.

Another important subject for further research is the case of a fully sampled aperture. Appropriate weighting can dramatically reduce the PSF side-lobe level resulting from a complete, and even partially overlapping, filling of the visibility plane. Such a weighting will inevitably increase the thermal noise level in a final image, but it reduces side-lobe noise considerably, which could result in lower total image noise. Even more importantly, it would minimize the amount of processing, and thus the required size of the imaging platform.

Comparing the subtraction of 20 sources for a low side-lobe configuration with the subtraction of about 100 sources for the LOFAR low band array, suggests a processing platform that is a factor 2.5 larger than the minimum one that is needed for a different array design. For the subtraction of up to 1000 sources, as is expected for the high band, this factor is as large as 25, requiring a post-correlation imaging platform of a size comparable to the correlation platform, to realize continuum imaging in real time for the Dutch LOFAR array.

However, full FoV imaging with the 10 times longer baselines of the European LOFAR configuration could require a processing platform for image forming that is a factor 100 larger than would be required for correlation. Such a platform will not be available in the coming few years, which means that only part of the observed FoV can be processed in practice. Especially the proposed fast faceting approach allows selecting those parts from the total FoV that provide the calibrations sources necessary for imaging a limited set of astronomically relevant objects.

We have identified wide-field high-resolution continuum imaging, which is the main application of a low frequency array, as an application that could drive the processing requirements for image-forming beyond what can be afforded when only a reasonable fraction of the total system cost of an aperture synthesis array is assigned to post-correlation processing.

Finally, we conclude by emphasizing that observing, where multi-direction self-calibration has to provide interpolated calibration, needs stations that satisfy minimum requirements. The first requirement is that the station has sufficient sensitivity to observe at least about 5 sources in its beam that can be used for second order interpolation. Especially when phase errors are induced by traveling ionospheric disturbances, we need sufficiently dense sampling of these structures requiring a beam width of about 4 degrees. In addition, it requires sufficient stations in an appropriate configuration to push the side-lobe noise below the ultra-low thermal noise that is aimed for after repeating many observations. For instance for imaging structures in the Universe that belong to the Epoch of Reionization (EoR) with the large low frequency array to be built as part of the SKA.

Samenvatting

Een belangrijke vraag bij een dissertatie is zijn classificatie. De dikte van dit boek plaatst het ergens tussen een leerboek en een verzameling publicaties gericht op een specialistisch publiek. Beeldvorming met een groot beeldveld is het hoofd onderwerp van de dissertatie. De centrale vraag daarbij is waardoor de verwerking van grote hoeveelheden interferometrische data fundamenteel bepaald wordt en hoe daarmee om te gaan. De niet-specialist wordt in een stapsgewijze uitleg geleid langs de benaderingen die nodig zijn voor efficiënte verwerking, culminerend in de schaal-wetten die het gebied van groothoek en breedspectrum beeldvorming beheersen.

Het ontwerp van een nieuw instrument brengt informatie samen uit vele disciplines. Het begint met een analyse van vergelijkbare instrumenten, in het bijzonder *hoe* en *waarom* ze werken. (Het *wat* is de derde dimensie van het kennis-volume, en komt later aan bod). Deze dissertatie laat zien *dat* een synthese radio telescoop zoals LOFAR kan worden gekalibreerd en vooral *waarom*. Dit vereist in de eerste plaats dat de stations bestaan uit een voldoende aantal element antennes, en bovendien dat de synthese *array* bestaat uit een voldoende aantal stations.

Het geheel van beschikbare wetenschappelijke kennis lijkt op een Zwitserse kaas; het heeft wel een wetenschappelijk verdedigbare structuur, maar bevat ook vele gaten. Het grootste deel van het bestreken volume bevat weinig materiaal dat geschikt is als startpunt voor systeem ontwerp. Neem bijvoorbeeld het ontwerp paradigma dat stelt dat alle onderdelen van een groot systeem gebaseerd moeten zijn op bewezen technologie. Dit is een recept voor vertraging van technische vooruitgang tot een tempo dat wordt bepaald door het voortschrijden van begrip en wacht op resultaten van deel implementaties. Hoewel het volgen van dit paradigma indekt tegen vertraging tijdens de zorgvuldig geplande bouw van een instrument, worden de struikelblokken die echte aandacht nodig hebben pas gevonden wanneer men er daadwerkelijk over struikelt. In werkelijkheid zijn de meeste dingen die echt werken helemaal nog niet volledig begrepen. Hoewel de wetenschap vele analytische gereedschappen heeft, ontbeert het de synthetische gereedschappen die nodig zijn voor constructie. Een eenvoudig voorbeeld is een slager, die precies weet hoe hij een varken moet ontleden in zijn verschillende onderdelen, maar niet in staat is om deze weer samen te voegen tot een compleet dier, laat staan een levend dier. In leven zijn is de essentie van iets dat werkt. De enig bekende manier om leven te creëren is vanuit het leven zelf, n.l. door levende elementen te nemen en die op organische wijze te laten samen samengroeien, terwijl ze hun weg zoeken langs en over de struikelblokken heen, op weg naar een niet erg goed gedefinieerde eindbestemming.

Dit organische synthese proces is weerspiegeld in deze dissertatie. Er worden over een breed front grote stappen gezet waar wetenschappelijk bewijs en volledig begrip voorhanden zijn, maar veel kleinere stappen in die gebieden waar nieuwe bruggen moeten worden gevonden om ons met minder inspanning naar de verlangde afbeeldingen te leiden. Hoewel het focus ligt op *waarom* laagfrequent radio synthese überhaupt kan werken is ook het *hoe* gepresenteerd om een nieuwe generatie gerust te stellen dat het beoogde doel inderdaad onder handbereik ligt. Veel details zijn nog niet ingevuld, zodat het riskant is om door jeugdig elan gedreven vooruit te stormen. Maar het nodigt uit om aandachtig voorwaarts te gaan, met een open oog voor problemen die onderweg moeten (en kunnen) worden opgelost.

Om die reden moet een systeem ontwerp beginnen met de identificatie van de fundamentele beperkingen, en leren hoe daarmee wordt omgegaan in vergelijkbare systemen. Praktische oplossingen worden meestal niet ingegeven door fundamentele overwegingen maar door problemen die worden veroorzaakt door voorbarige keuzes die zijn gemaakt in het conceptueel ontwerp. Deze keuzes zijn “voorbarig” omdat ze worden ingegeven door de kennis en technologie die op dat moment beschikbaar zijn, maar die achterhaald kunnen zijn tegen de tijd dat het systeem wordt opgeleverd.

Nieuwe elementen in het conceptueel ontwerp voor LOFAR

In de jaren negentig hebben 74 MHz waarnemingen met de VLA aangetoond dat de gevoeligheid onvoldoende was voor de zelf-kalibratie die nodig is om te corrigeren voor ionosferische verstoringen voor basislijnen tot een lengte van 30 km. Een nog groter probleem was het grote gezichtsveld van de schotel antennes die met hun diameter van 25 m groot lijken, maar gemeten in golflengtes te klein zijn. Om toch een bruikbaar plaatje kunnen maken zijn aanpassingen gemaakt in bestaande software pakketten, die waren ontwikkeld en geoptimaliseerd voor veel smallere antenna bundels bij kortere golflengtes.

De meest belangrijke conclusie hieruit voor toekomstige ontwerpen was de noodzaak voor voldoende gevoeligheid om de heldere bronnen in het veld van de stationsbundels te kunnen meten m.b.v. multi-directionele zelf-kalibratie. Dit vertaalde zich direct in de eis van grotere stations met voldoende gevoeligheid.

Rond de eeuwwisseling kwam digitale signaal verwerkings apparatuur beschikbaar die *phased-array* stations van veel meer dan 25 m diameter betaalbaar maakte voor de astronomie. Deze hebben, vanwege hun elektronisch gestuurde bundels, bovendien geen mechanisch volg-mechanisme nodig om astronomische objecten waar te nemen vanaf een draaiende Aarde. Nog belangrijker was dat de bekende wet van Moore voorspelde dat de verwerkingscapaciteit van digitale processors elke drie jaar zou verviervoudigen, waardoor de benodigde dataverwerking betaalbaar zou worden na 2003. Dit vertrouwen in de tijdige beschikbaarheid van

componenten en platforms was de basis van het concept ontwerp voor LOFAR, dat in 1999 werd gepresenteerd.

Uitgaande van voorlopige specificaties van de industrie voor toekomstige componenten en data verwerkings platforms kon het eigenlijke ontwerp van digitale ontvanger systemen en van kalibratie en beeldvormings software nu direct beginnen. Een meer uitvoerig overzicht van alle vernieuwende aspecten in LOFAR wordt gegeven in hoofdstuk 2. De meest belangrijke doorbraak voor de realisering van LOFAR was de aankondiging van goedkope gig bit transceiver technologie voor het data transport over glasvezels. Dit maakte het betaalbaar om stations te overwegen op locaties die honderden kilometers zijn verwijderd van de centrale data verwerking.

Het LOFAR teststation kwam gereed in 2003. Het maakte gebruik van de eerste generatie van nieuwe componenten; breedband korte dipool antennes, digitale ontvangers, kruis-correlatie op een cluster van processors en zelf-kalibratie op stations-niveau. Het vlakke antenne station was in staat om het gehele hemel-halfrond af te beelden, waarmee de precieze richting bepaald kon worden van de verschillende signaal bronnen en bewegende bronnen in een *snapshot* plaatje. De combinatie van meerdere snapshots leverde een hemelkaart die zelfs groter was dan een halfrond.

Efficiënte beeldvormings methodes

Deze pioniersarbeid gaf de richting aan voor analyse van de inherente beperkingen van 2-D Fourier beeldvorming, zoals behandeld in hoofdstuk 3. Deze analyse heeft geresulteerd in een voorstel voor twee nieuwe beeldvormings methodes, gebaseerd op een nieuwe combinatie van bestaande technieken om grootbeeld kaarten te maken met arrays waarin de interferometers niet co-planair zijn, d.w.z. niet in een vlak liggen.

De zogeheten *facet*-techniek maakt een groot aantal kleine plaatjes van delen (facetten) van het totale gezichtsveld door dezelfde data voor elk facet opnieuw te verwerken. De zogeheten *W-projectie*-methode maakt gebruik van een complexe quasi-convolutie om de data van niet-co-planaire basislijnen te corrigeren alvorens de noodzakelijke Fourier transformatie uit te voeren.

Beide methodes vereisen teveel processing om van praktisch nut te zijn voor LOFAR. Dit was aanleiding voor een fundamentele analyse van het gehele beeldvormings proces. Deze analyse bracht aan het licht dat beide methodes een beperkt kijkveld hebben als gevolg van de *extrinsieke* non-planarity bepaald door de projectie van de basislijnen op de kijkrichting. Deze grote en variërende projecties zijn het gevolg van de aardrotatie als een telescoop een punt aan de

hemel volgt. Een belangrijk aspect van de twee genoemde methodes is het gemak om te corrigeren voor rotatie van de basislijnen.

Voor een groot beeldveld is het daarom aantrekkelijker om uit te gaan van een methode die gebruik maakt van de veel kleinere *intrinsieke* non-planarity van een array, die het gevolg is van plaatsing van stations op een bolvormige aardoppervlak.

Een analyse van de complexe quasi-convolutie correctie methode heeft de precieze relatie tussen de mate van non-planarity en de grootte van het gezichtsveld laten zien. Dit wees de weg naar twee alternatieve methodes voor meer efficiënte beeldvorming. Beiden vereisen dezelfde minimale hoeveelheid quasi-convolutie bewerkingen.

De eerste methode volgt de conventionele benadering met extrinsieke non-planarity, maar maakt gebruik van iets grotere facetten dan gebruikelijk door slechts een beperkte quasi-convolutie correctie toe te passen. Deze methode is speciaal aantrekkelijk voor basislijnen die langer zijn dan een paar honderd km en minimaliseert de hoeveelheid benodigde bewerkingen m.b.v. een nieuwe snelle facetteringstechniek. Deze nieuwe techniek is effectief voor breedband waarnemingen zoals met LOFAR en is gebaseerd op een *butterfly* methode zoals wordt gebruikt in de Fast Fourier Transform. De totale hoeveelheid interferometer data wordt verdeeld over een aantal deelverzamelingen waarbij het totale data volume gelijk blijft. De essentie is dat elke kleine deelverzameling slechts een kleine Fourier transformatie vereist. Daardoor is de totale data verwerking voor alle facetbeelden samen gelijk aan de verwerking voor een enkel groot beeld, met het verschil dat de vervormingen door (intrinsieke en extrinsieke) non-planarity bijna volledig zijn gecorrigeerd.

De tweede methode is gebaseerd op individuele snapshotbeelden (beeld na korte waarneemperiode van een paar minuten). Aangezien de basislijnen voor een snapshotbeeld bij benadering in een vlak liggen heeft een 2-D Fourier transformatie een voldoende groot beeldveld. Hierbij wordt dus de grote extrinsieke non-planarity vermeden, die het gevolg is van projectie effecten. We hebben alleen te maken met de veel kleiner intrinsieke non-planarity, die wordt veroorzaakt door de kromming van het aardoppervlak waarop de stations staan. Een belangrijk bijkomend voordeel van deze methode is het directe inzicht in de wijze waarop beeldfouten tot stand komen als functie van positie aan de hemel. Deze fouten zijn niet alleen het gevolg van non-planarity, maar ook van effecten die samenhangen met de schijnbare rotatie van de hemel.

Een beperking van de methode is dat de waarneemtijd van een snapshotbeeld beperkt is tot ongeveer 10 minuten voor waarnemingen met stations tot 90 km van het centrum van de array. Een simpele eerste-orde correctie voor de rotatie tijdens deze periode is dan voldoende, omdat de residuale fouten dan van dezelfde orde zijn als die veroorzaakt worden door intrinsieke non-planarity

De individuele snapshotbeelden hebben nog een extra bewerking nodig voordat ze kunnen worden gecombineerd tot het uiteindelijke beeld van de hemel. Een schaal correctie is nodig omdat elk snapshotbeeld een andere projectie van de hemel is, terwijl de rotatie over het beeldveld verloopt. Naast deze positie correcties per 10 min. moeten ook intensiteits correcties uitgevoerd worden. Hierbij worden steeds 4 snapshotbeelden voor de vier polarisatie signalen gecombineerd om een plaatje voor elk van de 4 Stokes parameters te krijgen, die gecorrigeerd zijn voor de polarisatie veroorzaakt door de stations bundels en voor de parallaxische rotatie van de hemel polarisatie. Dit zijn correcties per image pixel die langzaam verlopen over het beeldveld.

Een synthese waarneming langer dan 10 min. heeft dus een aantal grote Fourier transformaties nodig, die de processing niet domineren voor continuüm waarnemingen met LOFAR en andere arrays met vergelijkbare of grotere aantallen stations.

De belangrijkste eigenschap van beide methodes is dat de benodigde computer verwerkingscapaciteit evenredig is met het aantal resolutie elementen in het verwerkte totale beeld. De implementatie van de voorgestelde methoden voor de imaging pakketten voor LOFAR is inmiddels gevorderd en zal invloed hebben op het uiteindelijk ontwerp voor een nog groter instrument zoals SKA.

System design en system engineering

De Square Kilometer Array (SKA) zal worden ontwikkeld en gebouwd in een aantal stadia. De technische uitwerking van het voorgestelde instrument vereist "bewezen" technologie om de omvang van geschikte platforms voor signaal en data verwerking te kunnen afschatten. Performance schaling van bestaande imaging pakketten leidt voor LOFAR tot processing platforms voor imaging, die veel groter zijn dan het platform voor de correlaties. Dit heeft tot de vraag geleid of deze pakketten de juiste optimalisatie hebben voor de veel grotere datastromen, of dat het imaging algoritme niet de optimale keus is voor arrays zoals LOFAR. In feite stellen we de vraag naar de fundamentele schalingswetten voor groot veld afbeelding op lage frequenties.

Systeem design houdt zich bezig met het zodanig combineren van onafhankelijke deelsystemen dat een gesteld doel wordt bereikt met minimale kosten. Systeem ontwerp voor wetenschappelijk onderzoek begint met een gegeven budget en streeft naar een maximale opbrengst van investering. Het is gebruikelijk dat wetenschappers de doelen stellen, waarna ingenieurs het instrument ontwerpen en de kosten uitrekenen. Hoewel het gebruik van uitsluitend bewezen technologie kan leiden tot voorspelbare kosten en tijdschalen, is er een gevaar dat het instrument achterhaald is bij oplevering.

System engineering houdt zich bezig met het opsplitsen van een groot systeem in aan aantal deelsystemen, zodat die onafhankelijk van elkaar in meer detail ontworpen kunnen worden. Dit is van belang om met meerdere teams parallel te kunnen werken en om teams met ingenieurs met verschillende specialisatie in te schakelen. Binnen het elektronica vakgebied onderscheiden we weer antenne, ontvanger en digitale ingenieurs en zijn duidelijke afspraken nodig hoe de signalen van het ene domein naar het andere domein over gaan. Tussen antenne en ontvanger zit veelal een transmissielijn. Een gebruikelijke afspraak is dat de antenne ingenieurs een antenne matchen aan een transmissie lijn en dat ontvanger ingenieurs de transmissie lijn matchen aan de lage ruis transistors.

System design houdt zich juist bezig met het definiëren van dergelijke overgangen, waardoor niet alleen grote besparingen mogelijk zijn, maar heel andere mogelijkheden ontstaan.

Een treffend voorbeeld van het opnieuw doordenken van oude waarheden is de korte dipool (i.e. veel korter dan de waarneem golflengte), die door ervaren antenne ingenieurs wordt beschouwd als een smalbandige component. Maar als we ons realiseren dat, voor onze toepassing, we alleen zijn geïnteresseerd in het ontvangen van signalen, en niet in uitzenden, kunnen we de beperkende eis van vermogens aanpassing deels laten vallen. In dit regiem kan een korte dipool breedbandig zijn, zelfs over twee octaven (een factor 4) in het LOFAR frequentie bereik.

Voor laagfrequente waarnemingen beantwoordt een station met korte dipolen aan de gestelde eisen: Een gegeven budget voor het totale aantal antenne elementen bepaalt de gevoeligheid van het systeem, waarna de antennes op verschillende manieren kunnen worden verdeeld over een aantal stations zonder de totale kosten significant te beïnvloeden. Een groot aantal kleine stations geeft een goede apertuur bedekking, terwijl een array met minder maar grotere stations minder databewerking nodig heeft. In het laatste geval heeft elk station meer bundels nodig om hetzelfde deel van de hemel te bestrijken, zodat het totale aantal ontvanger ketens niet verminderd is.

De beste keuze voor een configuratie van stations hangt af van de toepassing. Deze dissertatie concentreert zich op de meest gevoelige toepassing, die bovendien de meeste data-bewerking vereist: groothoekwaarnemingen met een grote relatieve bandbreedte.

De meest belangrijke vragen voor het systeem ontwerp zijn (I) hoe de configuratie en het aantal stations de zijlus ruis bepalen en de vereiste dataverwerking, en (II) hoe de grootte van de stations de kwaliteit van de zelf-kalibratie beïnvloedt en extra ruis introduceert. Hoofdstuk 3 bespreekt de minimum hoeveelheid dataverwerking die nodig is voor Fourier beeldvorming, en hoofdstuk 4 concludeert dat de minimale stations grootte wordt bepaald door de schaal van de ionosferische verstoringen.

Hoofdstuk 5 beschouwt hoe artefacten, die worden veroorzaakt door imperfecte kalibratie en door een beperkt aantal stations de gevoeligheid reduceren. Het is de gevoeligheid die voor het grootste deel de kosten van een synthese telescoop bepaalt, zodat elke beperking daarvan tot een equivalente prijs omgerekend kan worden. De resultaten van deze drie hoofdstukken zijn in hoofdstuk 6 gecombineerd tot een aantal conclusies en aanbevelingen voor systeem ontwerp en verder onderzoek.

Array configuratie en zijluis

In elk afbeeldingsinstrument wordt de afbeelding van een puntvormige bron *geconvolveerd* (versmeerd) met een zogenaamde *point-spread-function* (*psf*). De vorm van deze *psf* wordt bepaald door de bemonstering van het apertuur vlak. In radio apertuur synthese is deze bemonstering relatief incompleet waardoor de *psf* relatief hoge zijlussen heeft, die zich uitstrekken over het gehele beeldvlak. Zwakkere radiobronnen, die vaak het interessantst zijn, “verdrinken” daardoor in de *psf* van de sterkste bronnen. Daarom moeten de sterkste bronnen eerst worden gevonden en heel precies afgetrokken van de gemeten data alvorens de overgebleven residuen te transformeren tot een afbeelding die de zwakkere bronnen laat zien. Voor breedband continuüm waarnemingen met LOFAR zal deze operatie de dataverwerking domineren als meer dan 20 heldere bronnen moeten worden afgetrokken. Dit onderwerp wordt uitgebreid besproken in hoofdstuk 3. Daarom is de vraag hoeveel bronnen moeten worden afgetrokken heel belangrijk in het systeem ontwerp en dus ook hoe dat aantal wordt beïnvloed door de array configuratie. Dit onderwerp wordt behandeld in hoofdstuk 5. Aangezien het thermische ruis niveau in een Fourier afbeelding wordt verhoogd met het gemiddelde van de *psf* zijlussen van alle bronnen in het veld, is het zaak dat deze zijlussen zo klein mogelijk zijn. Dit kan worden bereikt door betere bemonstering van het apertuur vlak. Bijvoorbeeld door het gebruik van meer stations, die bovendien zorgvuldig zijn gepositioneerd, en/of door het gebruik van een grotere relatieve bandbreedte, en/of door langer waar te nemen, terwijl de aarde (en dus het synthese array) roteert. Bovendien kunnen de zijlussen aanzienlijk worden verminderd door het vermenigvuldigen van de gemeten data met zorgvuldig gekozen gewichts-factoren tijdens de Fourier transformatie. (NB: De *psf* zijlussen worden ook beïnvloed door kalibratie-fouten, die hier buiten beschouwing worden gelaten).

De bemonstering van het apertuur vlak wordt bepaald door de auto-correlatie van de stations-verdeling. In hoofdstuk 5 wordt een vereenvoudigd configuratie model voor LOFAR geïntroduceerd, waarin de helft van de stations zijn geconcentreerd in een centrale cluster en de andere helft in ringen met exponentieel toenemende stralen. Een dergelijk configuratie kan worden geoptimaliseerd voor continuüm waarnemingen door met zijn drie karakteristieke parameters te spelen: de diameter van een station, de diameter van de centrale cluster en de diameter van het gehele

array. Het quotiënt van de stations diameter gedeeld door de array diameter geeft een karakteristieke relatieve bandbreedte t.o.v. de waarneem frequentie en een karakteristieke waarneem tijd, die een geschikte bemonstering definiëren voor een stukje apertuur met de diameter van een station.

Het rms zijlus niveau van de psf waarmee de bronnen in een typische snapshot afbeelding zijn geconvolveerd wordt in de eerste plaats bepaald door het aantal stations in het array. Dit niveau neemt af met de afstand tot het centrum van de psf. Deze afname is een gecompliceerde functie, maar ver van het centrum is hij evenredig met de afstand. Met de bovengenoemde karakteristieke relatieve bandbreedte en snapshot waarneemtijd begint deze afname op een afstand die gelijk is aan de halfwaarde straal van de stations bundel.

In een multi-frequency synthese waarneming kunnen we deze snapshot waarneemtijd en relatieve bandbreedte vergroten totdat ze overeen komen met het quotiënt van de stations diameter gedeeld door de gemiddelde afstand tussen de stations in de centrale cluster. De apertuur wordt dan bemonsterd door een patroon van clusters van onafhankelijke data, wat leidt tot een vermindering van de straal waar de psf zijlus afname begint. Een verdere vermindering van de snapshot waarneemtijd heeft weinig effect aangezien we de apertuur dienen te bedekken met meer onafhankelijke clusters en niet met meer data per cluster.

Het maximum aantal onafhankelijke clusters wordt bepaald door het quotiënt van de array diameter gedeeld door de cluster diameter. Het demonstreert het belang van het distribueren van extra waarneem bandbreedte en tijd op een wijze die leidt tot de vorming van onafhankelijke data clusters in de apertuur. De belangrijke conclusie is dat de zijlus ruis in een synthese waarneming voor korte tijd en smalle bandbreedte op dezelfde wijze afneemt als de thermische ruis, nl. met de bandbreedte en waarneemtijd. Voor langere tijd en grotere bandbreedte is dit niet meer het geval.

Evaluatie m.b.v. dit vereenvoudigde model van de rms zijlus ruis voor het LOFAR laagfrequent array laat zien dat de ruis bijdrage van de zijlussen van alle bronnen buiten de stations bundel minder is dan 3% van de thermische ruis en dus minder dan 0.05% bijdraagt aan de ruis in de afbeelding. De voornaamste reden voor deze geringe bijdrage is dat de zijlussen van de stationsbundel de buitenliggende bronnen effectief onderdrukken, m.u.v. een paar zeer heldere. Die laatste zijn echter net helder genoeg om ze te kunnen zelf-kalibreren en dus netjes af te trekken van de gemeten data.

Het is een eigenschap van de intensiteits verdeling van de bronnen aan de hemel, dat de bronnen zwakker dan de 4 sterkste aan het noordelijk hemel halfrond één tot twee orde groottes zwakker zijn en niet individueel afgetrokken hoeven worden.

Uitgaande van een typische psf zijlus verdeling moeten tenminste 100 bronnen in de stations bundel worden afgetrokken voor een 12 uur waarneming met een array

van 40 stations en 1% relatieve bandbreedte bij 35 MHz. Na aftrekking is de totale zijlus ruis van alle overblijvende bronnen minder dan 40% van de thermische ruis. Deze bijdrage verhoogt de thermische ruis in de afbeelding met maximaal 8% en kan worden verminderd door meer bronnen af te trekken. Deze evaluatie gaat uit van perfecte aftrekking, hetgeen alleen mogelijk is voor de paar helderste bronnen in de stations hoofdlus en de paar helderste in de rest van de hemel, die individueel zijn gezelf-kalibreerd.

De invloed van de configuratie op zelf-kalibratie

Traditionele zelf-kalibratie lost op voor slechts een enkele complexe fout per station en neemt aan dat deze geldig is voor het hele gezichtsveld. Gegeneraliseerde zelf-kalibratie houdt rekening met het feit dat sommige fouten afhankelijk zijn van de kijkrichting en moet dus meer parameters per station oplossen. De kracht van multi-directionele zelf-kalibratie wordt uitgelegd voor niet-specialisten in hoofdstuk 4. Daar komt onder andere aan de orde in hoeverre het mogelijk is om kalibratie parameters te interpoleren, die snel veranderen in tijd, frequentie of positie. Bij voorbeeld de kleinschalige ionosferische instabiliteit, die varieert op tijdschalen van een minuut. Zulke overwegingen leiden tot een minimum afmeting (en dus gevoeligheid en veld-grootte) van een station.

De afleiding van de stationsgrootte gaat uit van fundamentele principes en zet de kleinschalige verstoringen, die de tijdsafhankelijke ionosferische refractie beschrijven, in de context van de grootschalige ionosferische structuur. In principe veranderen deze grootschalige termen slechts langzaam en kunnen op zich worden opgelost met behulp van zelf-kalibratie. Scheiding van de snellere kleinschalige effecten is mogelijk door middeling over intervallen van de orde 10 minuten, een typische waarde voor een snapshotbeeld, die ook door andere beperkingen wordt vereist.

Het blijkt dat zelf-kalibratie slechts kan oplossen voor een beperkt aantal parameters per station. Dit aantal wordt fundamenteel beperkt door het aantal onafhankelijke basislijnen waar een station in deelneemt, maar in praktijk door de ruis in de gemeten data. Iteratieve algoritmes lossen op voor parameters per bron-richting, in afnemende volgorde van helderheid, tot een flux van ongeveer driemaal de ruis op de data. Deze effectieve ruis bevat een component die veroorzaakt wordt door alle andere bronnen die te zwak zijn om voor opgelost te worden en die dus mede het maximum aantal bronnen bepalen waarvoor kan worden opgelost.

We leiden een eerste-orde schatting af, gebaseerd op de daadwerkelijke bron-dichtheid als een functie van flux, en laten zien dat dit aantal ongeveer 10 is, onder de vereenvoudigende aanname, dat de ruisbijdrage van alle vervuilende bronnen gelijk is aan de thermische ruis. Bovendien is een relatieve bandbreedte van meer dan 10% vereist voor een snapshot dataset waarvan de waarneemtijd gelijk is aan

een ionosferische coherentie tijd. Alleen basislijnen langer dan een km kunnen gebruikt worden, om de invloed van vervuillende bronnen te minimaliseren.

Een extra beperking is dat ook parameters moeten worden opgelost voor ongeveer 4 zeer heldere bronnen buiten de stations bundel (het zogenaamde A-team: Cas A, Cygnus A, etc.), zodat ze met grote nauwkeurigheid kunnen worden afgetrokken. De belangrijke conclusie is dat er opgelost kan worden voor tenminste 5 bronnen (richtingen) per station, wat voldoende is om de fase fouten over de stationsbundels te modelleren, die veroorzaakt worden door middelgrote Travelling Ionospheric Disturbances (TID), mits de bundel kleiner is dan ongeveer 4 graden.

LOFAR heeft voldoende stations van verschillende afmetingen voor voldoende gevoeligheid op verschillende basislijnen wat zelf-kalibratie en hoge kwaliteit groothoek beeldvorming mogelijk maakt in de hoogfrequent band (115-230 MHz). Groothoek beeldvorming in de laagfrequent band (10-90 MHz) wordt gecompliceerd door relatief grote hoeken tussen geschikte zelf-kalibratie bronnen vanwege grote stations bundels en beperkte gevoeligheid. Daardoor zijn de geïnterpoleerde fases slechts een globale afspiegeling van de daadwerkelijke ionosferische TID's. De grootte van de residuele fase fouten neemt toe met waarneem golflengte en de hoek tussen de bronnen die voor zelf-kalibratie worden gebruikt en maakt hoge kwaliteit afbeelding slechts mogelijk voor een beperkt deel van de stations bundel.

Een belangrijke vraag is op welke wijze de flux, die door de fase fouten wordt uitgespreid, in de Fourier afbeelding terecht komt. En vooral of dit de thermische ruis vermeerderd tot boven het niveau, dat wordt bepaald door de ontvanger ruis, de helderheid van de hemel en de waarneemtijd.

Alle bronnen die gebruik maken van geïnterpoleerde kalibratie parameters worden aangetast door fouten, die toenemen met de afstand tot de zelf-kalibratie bronnen (waar wordt aangenomen dat de fouten nul zijn). Deze interpolatie fouten veroorzaken een vervorming van de psf waar elke bron mee is geconvolveerd en leiden dus ook tot extra ruis vanwege de fouten in de zijlussen van deze bronnen. Een eerste-orde schatting, gebruik makend van de fase-ruis in de geïnterpoleerde kalibratie parameters, suggereert dat de residuele ruis na aftrekking minder is dan 38% van de thermische ruis. Dit draagt dan maximaal 7% bij aan de thermische ruis in de afbeelding, maar dit kan niet verder worden gereduceerd. Dit is een generiek resultaat dat geldig is voor de ruis, die veroorzaakt wordt door alle bronnen, die niet afgetrokken konden worden m.b.v. hun eigen zelf-kalibratie parameters, maar gebruik moesten maken van geïnterpoleerde parameters.

Naast deze bijdrages door geïnterpoleerde zelf-kalibratie parameters zijn er fase fouten die worden veroorzaakt door Kolmogorov turbulentie in de ionosfeer. Ook deze fase fouten nemen toe met de afstand tot de zelf-kalibratie bronnen. Voor de brede bundel van de laagfrequent LOFAR stations, waar die afstand relatief groot is, kunnen de fase fouten per coherentie tijd een waarde bereiken van meer dan 0.7

radiaal per station. Voor kortere basislijnen liggen de stations voldoende dicht bij elkaar om kalibratie informatie te delen, zodat geïnterpoleerde fouten kleiner zijn. Het gevolg is dat, voor een geven richting in de stations bundel, de fase fouten groter kunnen zijn dan 1 radiaal per basislijn.

Voor zulke grote fase-fouten kan de vervorming van de psf niet langer worden beschreven als perturbaties van de nominale psf. Het blijkt dat de resulterende psf een zijlus verdeling heeft met dezelfde rms waarde. Een belangrijk aspect is dat de hoofdlus van de psf ook zal worden opgebroken, wat resulteert in een zogenaamd speckle patroon. In dat geval zal het middelen van meerdere gespeckelde snapshot beelden ernstige versmering veroorzaken van bronnen in een groot deel van het veld. Dit reduceert niet alleen de piek intensiteit van objecten in die gebieden, maar geeft ook extra zijlussen met de verstrooide flux van deze bronnen. Een eerste orde schatting van deze ruis bijdrage, die gedomineerd wordt door de lange basislijnen, is minder dan 45% van de thermische ruis. Deze bijdrage kan dus worden gereduceerd door de lange basislijnen te negeren, maar dit vermindert dan ook de resolutie van alle goed-afgebeelde bronnen dichterbij de zelf-kalibratie bronnen.

Tenslotte moeten de verschillende fout-bijdrages worden gecombineerd door hun gekwadrateerde rms waarden bij elkaar op te tellen. Dit geeft een toename van tenminste 7% in de beeldruis van de LOFAR hoogfrequent band waarnemingen en typisch 23% voor de laagfrequent band.

Ontwerp optimalisatie en schaling van data verwerking

Voor effectief systeem ontwerp zijn schaalwetten nodig die de optimale verdeling van kosten over subsystemen bepalen voor een systeem met een beeldkwaliteit en effectieve gevoeligheid die aansluit bij de nominale gevoeligheid, die bepaald wordt door het opvangend oppervlak en de signaal bandbreedte. We hebben laten zien dat de uiteindelijke ruis in een afbeelding niet alleen bepaald wordt door de grootte van het instrument, maar ook door zijn configuratie. De afgeleide relaties maken het mogelijk om de invloed van verschillende configuraties op de totale systeem performance te vergelijken.

Een te kleine stationsafmeting biedt een beperkte kalibratie nauwkeurigheid, die leidt tot extra ruis die niet kan worden gereduceerd. Het gebruik van te weinig stations veroorzaakt een hoog psf zijlus niveau dat extra data-bewerkingen nodig maakt voor het aftrekken van bronnen. Bijvoorbeeld, een 10% verlies aan gevoeligheid door onvoldoende data-bewerking kan worden gereduceerd tot 5% door een kostbare vergroting van het beeldvormings platform voor het aftrekken van extra bronnen, wat alleen effectief is voor continuüm waarnemingen. Als alternatief kan het totaal aantal stations worden vergroot met 5%, wat eveneens kostbaar is en niet nodig voor spectraallijn waarnemingen. Of het totale beeldveld kan worden vergroot door 10% meer bundels te vormen, maar dit helpt alleen voor survey toepassingen.

In hoofdstuk 3 komen we tot de conclusie dat de minimum hoeveelheid dataverwerking voor continuüm beeldvorming evenredig is met de veld-grootte, gemeten in resolutie elementen, nl. evenredig met het kwadraat van het quotiënt van de array diameter gedeeld door de stations diameter. Ook wordt aangetoond dat de dataverwerking wordt gedomineerd door het aftrekken van bronnen als het aantal meer dan 20 bedraagt. Hoofdstuk 6 combineert dit resultaat met de resultaten van hoofdstuk 5 en verschaft globale schaalwetten voor signaal en data verwerking die worden samengevat in de volgende paragrafen.

Phased-array technologie maakt het mogelijk om het totaal betaalbare opvangend oppervlak op flexibele wijze te distribueren over een aantal stations, die zelfs van verschillende grootte mogen zijn, terwijl de effectieve grootte van beeldveld kan worden gevarieerd door het vormen van meer bundels per station. Voor een gegeven stations grootte wordt het aantal stations bepaald door het beschikbare budget.

Als alternatief kan een configuratie met minder maar grotere stations in principe dezelfde gevoeligheid verschaffen en geeft met meer bundels per station dezelfde totale beeldveld grootte. Hoewel de totale bandbreedte van alle signalen aan de input van de correlator gelijk is, vermindert de benodigde data-verwerkings capaciteit lineair met het aantal stations. Minder voor de hand liggend is dat de output data stroom voor continuüm waarnemingen op dezelfde wijze vermindert, wat aantrekkelijk is voor een processing platform, dat de data moet verwerken in realtime. Vanuit het standpunt van het ontwerp van het correlator platform lijkt de veel-bundel oplossing te prefereren.

Onze analyse heeft echter aangetoond dat een configuratie met minder stations minder gemeten data oplevert, wat leidt tot verminderde beeldkwaliteit. Hoewel de stations bundel smaller is, moet tenminste het zelfde aantal bronnen worden afgetrokken. Hetzelfde aantal bronnen in een smallere bundel betekent slechts dat zwakkere bronnen moeten worden afgetrokken om hetzelfde niveau van beeld-ruis te bereiken als met een configuratie met meer en kleinere stations. Deze stations hebben een bredere bundel die minder gevoelig is maar evenveel zelf-kalibratie bronnen detecteert. Het gevolg daarvan is dat de afstand tussen de zelf-kalibratie bronnen toeneemt en de kalibratie kwaliteit na interpolatie vermindert. Dit leidt eveneens tot extra ruis in een kaart.

De processing voor het aftrekken van bronnen is evenredig met het aantal basislijnen en domineert de beeldvorming van continuüm bronnen. In een configuratie met grote stations, en dus meer bundels moeten, op z'n minst evenredig meer bronnen afgetrokken worden. Aangezien het aantal basislijnen sterker gereduceerd is, neemt de totale processing voor continuüm beeldvorming af, net zoals de processing voor correlatie van alle telescoop signalen.

We hebben voor de hoogfrequent band van LOFAR aangetoond dat het aantal bronnen dat moet worden afgetrokken niet lineair maar progressief toeneemt als de fluxen van die bronnen in de buurt komen van de thermische ruis. In dat geval wordt het voordeel van het gebruik van minder maar grotere stations twijfelachtig.

Meer studie is nodig aangaande het minimum aantal benodigde stations

Het lijkt aantrekkelijk om te beginnen met een array met relatief weinig maar grote stations, waarvan het effectieve beeldveld in een later stadium kan worden vergroot door het vormen van meer bundels d.m.v. additionele apparatuur voor signaalverwerking. Voor continuüm waarnemingen geeft dit een zijlus ruis die wordt bepaald door het aantal stations en hun configuratie. Deze kan ruim boven de thermische ruis uitkomen, in het bijzonder wanneer de waarnemingen vele malen worden herhaald. De zijlus ruis is identiek is in elke waarneming en zal het uiteindelijke ruisnivo bepalen. Deze zijlus ruis kan alleen worden gereduceerd door meer stations toe te voegen aan de configuratie.

Optimalisatie van de configuratie van een synthese array met phased array telescopen is dus afhankelijk van het uiteindelijke gevoeligheids niveau dat bereikt moet worden.

Daarom is een belangrijk onderwerp voor verdere analyse de vraag of de extra bronnen, die afgetrokken moeten worden, in het geval van een veelheid van nauwe bundels überhaupt kunnen worden geïdentificeerd, aangezien ze veel dichter bij de ruisvloer liggen. Als dat niet het geval is moet een hogere ruis-vloer worden geaccepteerd voor continuüm waarnemingen, die niet kan worden verlaagd als in een later stadium meer dataverwerking beschikbaar komt.

Een ander belangrijk onderwerp voor verder onderzoek is de situatie van een volledig bemonsterde apertuur. Het verstandig gebruik van gewichten kan dan het psf zijlus niveau drastisch reduceren. Weliswaar zullen deze gewichten onvermijdelijk het thermisch ruisniveau verhogen in de uiteindelijke afbeelding, maar aangezien ze tegelijkertijd de zijlus ruis reduceren kan het totaal gunstig zijn. Zelfs belangrijker, het zou de totaal benodigde hoeveelheid dataverwerking minimaliseren en dus de afmetingen van het benodigde beeld-vormings platform.

De vergelijking van het aftrekken van 20 bronnen voor een configuratie met lage zijlussen met het aftrekken van 100 bronnen voor het LOFAR laagfrequent array suggereert een processing platform dat 2.5 maal groter is dan het minimum dat nodig is voor een ander array ontwerp. Voor het aftrekken van 1000 bronnen, zoals verwacht voor de hoogfrequente array, is dit een factor 25, wat een beeld-vormings platform vereist dat even groot is als het correlatie platform.

Sterker nog, groothoek beeldvorming met de 10 maal langere basislijnen van de Europese LOFAR configuratie, kan een beeld-vormings platform nodig hebben dat 100 keer groter is dan het correlatie platform. Zulk een platform zal niet beschikbaar zijn in de komende paar jaar, hetgeen betekent dat slechts een deel van het beeldveld kan worden verwerkt. Met name de voorgestelde snelle beeldvormings methode op basis van facets maakt het mogelijk om die gedeeltes van het beeldveld te selecteren die de kalibratie bronnen bevatten die nodig zijn voor het afbeelden van een beperkt aantal astronomisch relevante velden.

We eindigen met het benadrukken dat waarnemingen waarvoor geïnterpoleerde kalibratie parameters nodig zijn, verschaft door multi-directionele zelf-kalibratie, stations nodig hebben die voldoen aan zekere minimum eisen. De eerste eis is dat een station voldoende gevoelig is, zodat tenminste 5 bronnen in de bundel kunnen worden gebruikt voor tweede orde interpolatie. Vooral als fase fouten worden veroorzaakt door Travelling Ionospheric Disturbances (TID) moeten deze structuren voldoende worden bemonsterd, wat een bundel van ongeveer 4 graden vereist. Bovendien zijn er voldoende stations nodig geplaatst in een configuratie met minimaal psf zijlus niveau. De resulterende zijlus ruis moet minder zijn dan de ultra-lage thermische ruis, die mogelijk is door het vele malen herhalen van dezelfde waarnemingen. Bijvoorbeeld voor het afbeelden van structuur in de Epoch of Reionisation (EoR) met het grote laagfrequent array dat gebouwd gaat worden als deel van de SKA.

Dankwoord

Mijn dank gaat uit naar een aantal personen zonder wier inbreng dit proefschrift niet tot stand gekomen zou zijn.

In 2007 heb ik op aanraden van prof. Ir. A. van Ardenne, toenmalig directeur van de divisie Emerging Technologies van ASTRON en mijn directe leidinggevende, besloten om een aantal relevante aspecten van het ontwerp van LOFAR vast te leggen als referentie voor een nieuwe generatie ontwerpers. Dit is nu gebeurd en wel in de vorm van een proefschrift, deels op basis van mijn bijdragen vastgelegd in publicaties. Tevens was duidelijk geworden dat het vigerende concept voor beeldvorming met LOFAR niet toestond een afschatting te maken van de benodigde rekenkracht daarvoor. Met het oog op de plannen voor een nog grotere synthese radio telescoop, eveneens gebaseerd op phased array antenne stations, leek een nadere studie van de imaging aspecten een gewenste aanvulling voor het proefschrift. Op basis van deze concept opzet heb ik twee hoogleraren benaderd met de vraag om het proces naar een promotie te begeleiden.

Prof. Dr. W.N. Brouw, een van de grootste experts op het gebied van synthese imaging, heb ik gevraagd als promotor en begeleider op te treden.

Wim, jij noch ik kon toen bevroeden wat de consequenties daarvan zouden worden. Als eerst verantwoordelijke voor de wetenschappelijke integriteit heb jij in belangrijke mate bijgedragen aan ruim een verdubbeling van de omvang van de geplande dissertatie.

Wim, ik wil je in het bijzonder bedanken voor je niet aflatende inzet om precisie en correctheid tot stand te brengen en steeds weer de vinger te leggen op zwakke plekken in opeenvolgende iteraties van het manuscript.

Prof. Dr. H.R. Butcher, toenmalig directeur van ASTRON, heeft het programma gestart dat tot de realisatie van LOFAR heeft geleid. Dit programma heeft mij de kans gegeven om de verschillende concepten voor een systeem ontwerp verder uit te werken, samen met collega's van ASTRON, MIT en NRL. Dit heeft tot concrete demonstratie geleid van alle voorgestelde nieuwe technologieën en vormt de basis voor het eerste deel van de dissertatie.

Harvey, bedankt voor je langjarige ondersteuning van het promotie traject, speciaal voor jouw accentuering van de vele bijdragen, die soms verdrinken in alle details die essentieel zijn om duidelijk te maken waarom een systeem ook écht kan werken.

Het proefschrift is voor mij de bekroning van 40 jaar samenwerking met vele collega's bij ASTRON, om instrumentele projecten te laten werken tot aan de grens van hun potentiële mogelijkheden. Ik wil al deze collega's bedanken voor de plezierige samenwerking die ik ervaren heb, maar ik wil een paar personen in het bijzonder bedanken voor het tot stand komen van dit proefschrift.

Stefan (Wijnholds), een speciaal dankwoord voor jou. Na een periode waarin ik als jouw begeleider optrad, hebben we samen aan een groot aantal aspecten van phased arrays gewerkt, en in het bijzonder aan LOFAR. Het focus lag daarbij op de haalbaarheid van verschillende doelstellingen en de vereiste voorwaarden daarvoor. Je hebt je promotie sneller afgerond en op jouw beurt mij gesteund om de hoofdlijnen scherp in beeld te houden en als paranimf bij te staan.

Jan (Noordam), zonder jouw inspirerend optreden en twee decennia discussie over calibratie, was er geen basis geweest voor het concept ontwerp voor LOFAR. Het calibratie hoofdstuk is de weerslag van deze discussies en maakt een synthese van de verschillende concepten die voor LOFAR gebruikt kunnen worden. Extra dank voor je inzet als paranimf om op de valreep de leesbaarheid van een aantal secties te vergroten.

Ger (de Bruyn), als praktiserend observationeel astronoom en later als project scientist voor LOFAR, ben jij altijd mijn eerste aanspreekpunt geweest om nieuwe technische mogelijkheden te bediscussiëren op astronomische gebruiksmogelijkheden, veelal ver buiten je eigen specifieke onderzoeksgebieden. Bedankt voor de directe toegang tot je encyclopedische kennis en relevant waarneem materiaal waarvan ik bijna 30 jaar gebruik heb mogen maken.

Arnold (van Ardenne), je hebt het "Laboratorium" verbreed tot een dynamische R&D divisie waarin de basis is gelegd voor de grootschalige toepassing van phased array antenne stations in de radio astronomie. Ik wil jou bedanken voor de ruimte die je me al die jaren hebt gegeven om in een groot aantal projecten aan de wieg van nieuwe ontwikkelingen te kunnen staan en medewerkers daarbij op gang te helpen om potentiële mogelijkheden te onderzoeken en te ontwikkelen tot praktische realisaties.

Naast deze hoofdpersonen, wil ik een aantal collega's bedanken. Daarbij denk ik in de eerst plaats aan Johan (Hamaker), de kopman van het Hamaker-Bregman-Sault formalisme dat de basis vormt voor calibratie en imaging van polarisatie. Johan, bedankt voor het omzetten van mijn schetsen in nette tekeningen.

Dion (Kant) -the wizzard who makes it really work- je bent instrumenteel geweest in het realiseren van het Initial Test Station voor LOFAR met antennes, simulaties, ontvangers, digitale signaal processing op een cluster van computers, en besturing van het geheel. Samen met Stefan heb je de basis gelegd voor alle relevante demonstraties van de nieuwste phased array technieken.

Ten slotte, een aparte waardering voor Andre (Gunst), die als LOFAR system engineer open stond voor mijn vele suggesties. Vooral zijn opmerking "maak je geen zorgen, we maken er best iets moois van" heeft hij waar gemaakt en heeft mij de nodige rust gegeven deze dissertatie te schrijven.

**System Design and Wide-field Imaging Aspects
of Synthesis Array with Phased Array Stations**

To the next generation of SKA system designers

J.D.Bregman, Eext

Omslagontwerp en lay-out
O&D producties, William Oosterwijk, Eext
<http://www.design-your-thesis.nl>

Tekeningen
Johan Hamaker, Dwingeloo

Foto voorkant omslag
Top-Foto, Assen

Digitale bewerking omslag foto's
William Oosterwijk, Eext

Foto achterkant omslag
William Oosterwijk, Eext

Druk
CPI Koninklijke Wöhrman B.V. Zutphen

Copyright: J.D.Bregman, Eext 2012

ISBN: 978-90-367-5841-3

Electronische versie via Bibliotheek Rijksuniversiteit Groningen

ISBN: 978-90-367-5840-6

The author graduated in 1970 from the Technical University Delft, receiving a degree in applied physics. During his studies he was awarded a prize for innovative research in the field of “approximate Fast Fourier Transform processing”. This subject has played an important role in his further career, also in this thesis.

The dissertation is the culmination of 40 years at ASTRON. The first decade was spent as “instrument physicist” of the famous synthesis radio telescope at Westerbork (WSRT), bringing it to its full potential in automated mode.

The next decade was devoted to designing detector systems for optical telescopes, and on experiments in optical synthesis imaging, applying the highly successful self-calibration techniques from the radio domain.

Back to radio, to participate in upgrading the WSRT with cryogenic receiver systems, and developing new capabilities for the VLBI network in Europe. A sabbatical year in 1996 at Colorado University on antennas and phased arrays was followed by participation in the further development of these concepts for the Square Kilometre Array (SKA). One such concept evolved into the basic design for LOFAR.

Successful demonstration of the new technologies and techniques with the LOFAR initial test station in 2003 brought the prestigious “Veder Prize” of the Dutch Electronics and Radio Society (NERG).

The last decade was devoted to guiding the detailed design of LOFAR, by explaining the rationale behind all elements. This led to the present dissertation. Starting with an overview of all the relevant design elements, the focus is on efficient processing of the huge data volumes that are produced by the new generation of radio telescopes.

back picture: the author at a LOFAR low band station near his own backyard

This book presents a body of knowledge needed for design that bridges the gap between what is not explained in textbooks, but assumed by scientific papers.



These 4 simple antenna wires receive signals from beyond the visible universe, to help us choose between the many fanciful projections of mere human speculation.